

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-230572

(43)Date of publication of application : 16.08.2002

(51)Int.Cl. G06T 13/00

G06T 5/20

G06T 7/00

G06T 7/20

G06T 15/00

H04N 5/262

(21)Application number : 2001-002070 (71)Applicant : MONOLITH CO LTD

(22)Date of filing : 10.01.2001 (72)Inventor : SHINAGAWA YOSHIHISA
NAGASHIMA HIROKI
AKIYOSHI GIYOUZO
AKIYOSHI NOBUO

(30)Priority

Priority number : 2000360640

Priority date : 28.11.2000

Priority country : JP

(54) METHOD AND DEVICE FOR IMAGE EFFECT, METHOD AND DEVICE FOR

IMAGE INTERPOLATION

(57)Abstract:

PROBLEM TO BE SOLVED: To easily provide image effect of high quality with light data.

SOLUTION: An image inputting part 12 inputs a first image and a second image. A matching processor 14 performs pixel matching between those images, and a corresponding point on a lattice point of a mesh placed on the first image is obtained on the second image. A result is recorded as a corresponding point file F. An intermediate image generating part 18 generates an intermediate image between the first image and the second image on the basis of the corresponding point file F. Because the corresponding point file is created by only the lattice point, capacity of data is very small, and various image effect, animation compression, on-line goods introduction, and mounting on a portable apparatus are easily achieved.

LEGAL STATUS [Date of request for examination] 24.01.2001

[Date of sending the examiner's decision of rejection] 01.07.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration] withdrawal

[Date of final disposal for application] 22.08.2003

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The image effectiveness approach characterized by including the process which detects the point on the 2nd image corresponding to the lattice point of the mesh prepared on the 1st image by image matching, and the process which defines the terminal point polygon corresponding to the origin polygon which constitutes the mesh on the 1st image on the 2nd image based on the result.

[Claim 2] Said process to detect is a method according to claim 1 of performing matching count of a pixel unit between the 1st image and the 2nd image.

[Claim 3] Said process to detect is a method according to claim 2 of performing matching count of a pixel unit based on correspondence with the singular point which detected by performing two-dimensional retrieval about the 1st image, and the singular point which detected by performing two-dimensional retrieval about the 2nd image.

[Claim 4] Said process to detect is the approach according to claim 3 of acquiring the correspondence relation of the pixel unit in level with the final finest resolution, while succeeding to the matching count in resolution level which performs matching count of an image unit between the same resolution level, and is different in the result after forming the 1st image and the 2nd image into multiplex resolution by extracting said singular point, respectively.

[Claim 5] Said process to define is an approach given in either of claims 1-4 which output the correspondence relation between an origin polygon and a terminal point polygon to a corresponding-points file, relate this file with the 1st image and the 2nd image, and hold it.

[Claim 6] The approach according to claim 5 of including further the process which uses said corresponding-points file and generates the middle image of the 1st image and the 2nd image by interpolation count of a polygon unit.

[Claim 7] Said process to generate is the approach according to claim 6 of deriving the point inside the terminal point polygon corresponding to the point inside an origin polygon.

[Claim 8] The method according to claim 7 of dividing an origin polygon into a triangle and searching for the point inside a terminal point polygon in the internal division of the vector about three square shapes each.

[Claim 9] This matching processor is the image effectiveness equipment characterized by generating a corresponding-points file including the image input section which acquires the 1st image and the 2nd image, and the matching processor which calculates matching of the 1st image and the 2nd image by defining the terminal point polygon corresponding to the origin polygon which constitutes the mesh on the 1st image on the 2nd image.

[Claim 10] Said matching processor is equipment according to claim 9 which performs matching count of a pixel unit based on correspondence with the singular point which detected by performing two-dimensional retrieval about the 1st image, and the singular point which detected by performing two-dimensional retrieval about the 2nd image.

[Claim 11] Said matching processor is equipment according to claim 10 which acquires the correspondence relation of the pixel unit in level with the final finest resolution while succeeding to the matching count in resolution level which performs matching count of an image unit between the same resolution level, and is different in the result after forming the 1st image and the 2nd image into multiplex resolution by extracting said singular point, respectively.

[Claim 12] Equipment given in either of claims 9-11 which include further the communications department for sending out said corresponding-points file to the exterior.

[Claim 13] The image effectiveness equipment characterized by having described the physical relationship of the lattice point of a mesh when said corresponding-points file was taken on the 1st image, and the point on the 2nd image corresponding to it including the middle image generation section which acquires the 1st image, the 2nd image, and the corresponding-points file between them, and generates the middle image of these two images.

[Claim 14] Equipment according to claim 13 which contains further the display which adjusts the timing of said middle image, said 1st image, and said 2nd image, and displays them.

[Claim 15] The computer program characterized by making a computer perform the process which detects the point on the 2nd image corresponding to the lattice point of the mesh prepared on the 1st image by image matching, and the process which defines the terminal point polygon corresponding to the origin polygon which constitutes the mesh on the 1st image on the 2nd image based on the result.

[Claim 16] the image interpolation approach which characterizes by to include the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of the middle image, and offer the goods presentation function in electronic commerce.

[Claim 17] the image interpolation approach characterized by to include the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and

the 2nd image, and the process which are in charge of generation of the middle image, and offer the function about the image effectiveness.

[Claim 18] the image interpolation approach characterized by to include the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of the middle image, and offer the utility feature about generation of a dynamic image.

[Claim 19] the image interpolation approach characterized by to include the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which perform generation of the middle image under the function on condition of application to mobile computing devices.

[Claim 20] Said goods presentation function is an approach including the function which rotates and displays goods according to the directions from a user according to claim 16.

[Claim 21] Said goods presentation function is an approach including the function which displays the middle image which interpolated the goods image in the two directions and was obtained according to claim 16.

[Claim 22] The function about said image effectiveness is an approach including the function which selects two images as the 1st image and the 2nd image one by one from three or more images, and displays a middle image continuously according to claim 17.

[Claim 23] The function about said image effectiveness is an approach including the function which displays the middle image which interpolated the image set as the object of processing in the two directions, and was obtained according to claim 17.

[Claim 24] Said utility feature is an approach including the function to specify the number of sheets of the middle image which should be generated according to claim 18.

[Claim 25] The assignment about the number of sheets of said middle image which should be generated is an approach according to claim 24 acquired from a user through GUI.

[Claim 26] The function on condition of application to said mobile computing devices is the approach according to claim 19 of being the function which displays the middle image generated based on one side and said correspondence information on the 1st image or the 2nd image.

[Claim 27] The approach according to claim 26 by which streaming processing of one

side and said correspondence information on the 1st image or the 2nd image is carried out, and said middle image is generated.

[Claim 28] Are equipment which interpolates an image and the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image is acquired. the correspondence information -- the image interpolation equipment characterized by the middle image generation section concerned being in charge of generation of a middle image, and offering the goods presentation function in electronic commerce including the middle image generation section which generates the middle image of the 1st image and the 2nd image on a basis.

[Claim 29] Are equipment which interpolates an image and the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image is acquired. the correspondence information -- the image interpolation equipment characterized by the middle image generation section concerned being in charge of generation of a middle image, and offering the function about the image effectiveness including the middle image generation section which generates the middle image of the 1st image and the 2nd image on a basis.

[Claim 30] Are equipment which interpolates an image and the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image is acquired. the correspondence information -- the image interpolation equipment characterized by the middle image generation section concerned being in charge of generation of a middle image, and offering the utility feature about generation of a dynamic image including the middle image generation section which generates the middle image of the 1st image and the 2nd image on a basis.

[Claim 31] Are equipment which interpolates an image and the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image is acquired. the correspondence information -- the image interpolation equipment characterized by the middle image generation section concerned performing the display of a middle image under the function on condition of application to mobile computing devices including the middle image generation section which generates the middle image of the 1st image and the 2nd image on a basis.

[Claim 32] Said goods presentation function is equipment including the function which rotates and displays goods according to the directions from a user according to claim 28.

[Claim 33] Said goods presentation function is equipment including the function which displays the middle image which interpolated the goods image in the two directions and was obtained according to claim 28.

[Claim 34] The function about said image effectiveness is equipment including the function which selects two images as the 1st image and the 2nd image one by one from three or more images, and displays a middle image continuously according to claim 29.

[Claim 35] The function about said image effectiveness is equipment including the function which displays the middle image which interpolated the image set as the object of processing in the two directions, and was obtained according to claim 29.

[Claim 36] Said utility feature is equipment including the function to specify the number of sheets of the middle image which should be generated according to claim 30.

[Claim 37] The assignment about the number of sheets of said middle image which should be generated is equipment according to claim 36 acquired from a user through GUI.

[Claim 38] The function on condition of application to said mobile computing devices is equipment according to claim 31 which is the function which displays the middle image generated based on one side and said correspondence information on the 1st image or the 2nd image.

[Claim 39] Equipment according to claim 38 with which streaming processing of one side and said correspondence information on the 1st image or the 2nd image is carried out, and said middle image is generated.

[Claim 40] Equipment given in either of claims 28–39 which contain further the display which adjusts the timing of said middle image, said 1st image, and said 2nd image, and displays them.

[Claim 41] the computer program which characterizes by to make a computer perform the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of the middle image, and offer the goods presentation function in electronic commerce.

[Claim 42] the computer program which characterizes by to make a computer perform the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of the middle image, and offer the function about the image effectiveness.

[Claim 43] the computer program characterized by to make a computer perform the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which

constitute the mesh prepared on the 2nd image, the process which are alike also in the correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of the middle image, and offer the utility feature about generation of a dynamic image.

[Claim 44] The process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image, the computer program characterized by making a computer perform the process which is alike also in the correspondence information and generates the middle image of the 1st image and the 2nd image, and the process which performs generation of the middle image under the function on condition of application to mobile computing devices.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] Especially this invention relates to the digital image effectiveness technique about an image processing technique.

[0002]

[Description of the Prior Art] BS digital broadcasting greets the timing of initiation, and many users come to enjoy the service on the Internet from PC or a gestalt telephone, and the fence of broadcast and a communication link is collapsing quickly. in a broadband time, multimedia culture accomplishes true progress -- will come out and I will be -- it carries out, and conflicts in it and distribution of a reason animation serves as key technology. In the amount of the information which human being acquires from the external world, an image is said to be able to transmit 200 times as many contents as voice. An image does not remain in an amusement application but serves as a life of human being and a software infrastructure supporting culture widely. The place of activity of the image effectiveness technique also spreads by application of CG or an image processing technique as an image is digitized.

[0003]

[Problem(s) to be Solved by the Invention] Before, various image effectiveness is proposed and used. For example, it is walk-through, multiaspect image, special filtering, and false stereoscopic vision etc. To be accompanied by the high data compression effectiveness is desired so that such image effectiveness may be offered with higher quality and it may fill the request of a transmission system to coincidence, as the importance in the society of an image will increase further from now on. This invention is made to such a request and one of the purpose of that is in offer of the image

effectiveness technique in which comparatively light data realize generation of an animation. Another purpose of this invention is in offer of the image interpolation technique of realizing various imaging technology not only in the image effectiveness but digital one and the Internet age.

[0004]

[Means for Solving the Problem] The mode with this invention includes the process which detects the point on the 2nd image corresponding to the lattice point of the mesh prepared on the 1st image by image matching, and the process which defines the terminal point polygon corresponding to the origin polygon which constitutes the mesh on the 1st image on the 2nd image based on the result.

[0005] The process to detect may perform matching count of a pixel unit between the 1st image and the 2nd image. This count may follow all pixels and may follow the pixel relevant to the lattice point or the lattice point, and it.

[0006] The process to detect may perform matching count of a pixel unit based on correspondence with the singular point which detected by performing two-dimensional retrieval about the 1st image, and the singular point which detected by performing two-dimensional retrieval about the 2nd image. In this case, further, after forming the 1st image and the 2nd image into multiplex resolution by extracting the singular point, respectively, the process to detect performs matching count of an image unit between the same resolution level, and it may acquire the correspondence relation of the pixel unit in level with the final finest resolution, succeeding to the matching count in resolution level which is different in that result.

[0007] The process to define outputs the correspondence relation between an origin polygon and a terminal point polygon to a corresponding-points file, may relate this file with the 1st image and the 2nd image, and may hold it. Moreover, this invention may use a corresponding-points file and may also include further the process which generates the middle image of the 1st image and the 2nd image by interpolation count of a polygon unit.

[0008] These people are application of the technique (henceforth a "premise technique") previously proposed by patent No. 2927350, and are suitable for the matching approach using the above-mentioned singular point at said process to detect. However, the premise technique is not touching at all the description of this invention about the polygon decided by the lattice point or it. Since a kind of simplification technique of a polygon was introduced by this invention, size of said corresponding-points file can be made very small.

[0009] That is, if correspondence of those pixel units is described as it is when the 1st and 2nd image has a $n \times m$ pixel, respectively, the combination 2 ($n \times m$) will arise and a corresponding-points file will become the huge amount of data. However, the amount of data becomes less sharply by changing this into describing the correspondence relation of the lattice points, or the correspondence relation of polygons decided by

the lattice point although it is the same substantially.

[0010] On the other hand, if there are the 1st and 2nd image and this corresponding-points file, a middle image is generable by interpolation count. Therefore, this invention realizes morphing as used in the field of the image effectiveness. If the 1st and 2nd image is considered [both] to be a frame in an animation, this invention can be grasped as the compression technology of an animation instead of morphing. Therefore, if only the 1st and 2nd image and a corresponding-points file are held, an animation can be reproduced and big effectiveness will be demonstrated in transmission of an animation, storage, etc.

[0011] Another mode of this invention is related with the image effectiveness equipment. This matching processor generates a corresponding-points file including the image input section from which this equipment acquires the 1st image and the 2nd image, and the matching processor which calculates matching of the 1st image and the 2nd image by defining the terminal point polygon corresponding to the origin polygon which constitutes the mesh on the 1st image on the 2nd image.

[0012] A matching processor may perform matching count of a pixel unit based on correspondence with the singular point which detected by performing two-dimensional retrieval about the 1st image, and the singular point which detected by performing two-dimensional retrieval about the 2nd image. Moreover, after forming the 1st image and the 2nd image into multiplex resolution by extracting said singular point, respectively, matching count of an image unit is performed between the same resolution level, and the correspondence relation of the pixel unit in level with the final finest resolution may be acquired, succeeding to the matching count in resolution level which is different in the result.

[0013] Further, this equipment could acquire the 1st image, the 2nd image, and the corresponding-points file between them, and may describe the physical relationship of the lattice point of a mesh when said corresponding-points file was taken on the 1st image, and the point on the 2nd image corresponding to it including the middle image generation section which generates the middle image of these two images. Moreover, the display which adjusts the timing of a middle image, the 1st image, and the 2nd image, and displays them may also be included further. In addition, in a premise technique, this invention does not have an end as it is indispensable.

[0014] A mode is related with another pan of this invention at the image interpolation approach. this approach includes the process which acquires the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitute the mesh prepared on the 2nd image, the process which are alike also in that correspondence information and generate the middle image of the 1st image and the 2nd image, and the process which are in charge of generation of that middle image, and offer the goods presentation function in electronic commerce.

[0015] It may replace with the function and the function about the image effectiveness and the utility feature about generation of a dynamic image may be offered. Moreover, the display of a middle image may be performed under the function on condition of application to mobile computing devices.

[0016] A goods presentation function may also include the function which rotates and displays goods according to the directions from a user. Moreover, you may also include the function which displays the middle image which interpolated the goods image up and down with two directions, for example, right and left, and was obtained.

[0017] The function about the image effectiveness may also include the function which selects two images as the 1st image and the 2nd image one by one from three or more images, and displays a middle image continuously. For example, a panoptic image will be obtained, if between many discrete view images is interpolated one after another and a middle image is generated. If it displays in order as if it moved the view with time amount, a walk-through image will be obtained. The function about the image effectiveness may also include the function which displays the middle image which interpolated the image set as the object of processing in the two directions, and was obtained.

[0018] The utility feature about an animation may also include the function to specify the number of sheets of the middle image which should be generated. If many middle images are generated, a slow motion image is generable, for example. By specifying number of sheets, the way of setting up the reproduction speed of an animation interactively is opened. The assignment about the number of sheets of the middle image which should be generated may be acquired from a user through GUI.

[0019] The function on condition of application to mobile computing devices may be a function which displays the middle image generated based on one side and said correspondence information on the 1st image or the 2nd image. In that case, there is little capacity of the memory for setting an image temporary, and it ends. What is necessary is just to move the image element (for it to be made to only represent with a pixel below) of each pixel and others of the 1st image according to correspondence information, when only correspondence information and the 1st image generate a middle image. That is, although both the location of a pixel and a pixel value, i.e., a color, can usually be interpolated for correspondence information, this is simplified to processing of only migration of the location of a pixel. Of course, it is also possible to include the color information on the pixel of the 2nd image in the correspondence information itself, and both the location of a pixel and a color can interpolate good in that case.

[0020] Streaming processing of one side and said correspondence information on the 1st image or the 2nd image is carried out, and a middle image may be generated. The capacity of memory required for processing may be reduced also in this case, and it is advantageous to mounting to the terminal of a cellular phone and others.

[0021] Still more nearly another mode of this invention is equipment which interpolates an image, and the correspondence information on the origin polygon which constitutes the mesh prepared on the 1st image, and the terminal point polygon which constitutes the mesh prepared on the 2nd image is acquired. the correspondence information -- including the middle image generation section which generates the middle image of the 1st image and the 2nd image on a basis, the middle image generation section concerned is in charge of generation of a middle image, and offers the goods presentation function in electronic commerce. In addition, the function about the image effectiveness, the utility feature about generation of a dynamic image, and the function on condition of application to mobile computing devices may be offered.

[0022] This equipment may also contain the display which adjusts the timing of a middle image, the 1st image, and the 2nd image, and displays them further.

[0023] In addition, a part or the thing which replaced all, added or changed the expression into the computer program, the record medium, etc. is also effective in an expression as this invention between an approach and equipment in changing each above configuration and a process to arbitration.

[0024]

[Embodiment of the Invention] First, the image matching processing using the multiplex resolution singular point filter technique and it which are used with the gestalt of operation is explained in full detail as a "premise technique." These techniques are techniques in which these people have already acquired patent No. 2927350, and are suitable for combination with this invention. Since a mesh is prepared on an image and many pixels are represented with this invention by the lattice point, it is because the application effectiveness over the matching technique of a pixel unit like a premise technique from the first is high. However, an image matching technique employable with the gestalt of operation is not restricted to this.

[0025] The image interpolation technique and the image effectiveness technique which the premise technique was used are concretely explained after drawing 18.

[the background of a premise technique] -- matching of automatic matching of two images, i.e., an image field, and pixels is one of the most difficult and, important themes in a computer vision or computer graphics. For example, if matching can be taken between the images from a different view about a certain object, the image from other views is generable. If matching of a right eye image and a left eye image is calculable, the photogrammetry using a solid image is also possible. When matching of the model of the image of a face and the image of other faces is able to be taken, the part of characteristic faces, such as an eye, a nose, and opening, can be extracted. For example, when matching is correctly taken between the images of people's face and the face of a cat, morphing can be automated fully by generating the inside rate image of them automatically.

[0026] However, conventionally, people had to specify the corresponding points between two images one by one, and, generally they required the great activity man day. In order to solve this problem, many corresponding-points automatic detection approaches are proposed. For example, there is an idea of reducing the number of the candidates of corresponding points, by using an epipolar straight line. However, processing is very complicated even in such a case. In order to reduce complexity, it is assumed that the coordinate of each point of a left eye image is usually in the location almost same as a right eye image. However, if such constraint is prepared, it will become very difficult to take matching which fulfills the global description and the local description to coincidence.

[0027] In volume rendering, since the voxel is constituted, a series of cross-section images are used. In this case, conventionally, it is assumed that it corresponds to general with the pixel which has a pixel in an upper cross-section image in the same part of a downward cross-section image, and the pair of these pixels is used for interpolation count. Thus, in order to use a very simple approach, when a long distance and the cross-section configuration of an object are large and the distance between continuous cross sections changes, the object built by volume rendering tends to become not clear.

[0028] A stereophotogrammetry method etc. also has many matching algorithms using detection of an edge. However, since there are few corresponding points obtained as a result in this case, interpolation count of the value of disparity must be carried out in order to bury the gap between the corresponding points which were able to take matching. Generally, when the brightness of a pixel changes in the local window which they use, it is difficult for all edge detectors to judge whether this suggests existence of an edge truly. An edge detector is a high-pass filter altogether essentially, and a noise will also gather it to an edge and coincidence.

[0029] The optical flow is known as still more nearly another technique. When the image of two sheets is given, in an optical flow, a motion of the object in an image (rigid body) is detected. It is assumed in that case that the brightness of each pixel of an object does not change. In an optical flow, the motion vector (u, v) of each pixel is calculated with some additional conditions of the smoothness of vector field (u, v) . However, in an optical flow, the global correspondence relation between images is undetectable. It is only taking notice of a local change of the brightness of a pixel, and the error of a system becomes remarkable when the variation rate of an image is large.

[0030] In order to recognize the global structure of an image, many multiplex resolution filters have also been proposed. They are classified into a linearity filter and a nonlinear filter. Although there is wavelet as a former example, generally a linearity filter is not very useful to image matching. Because, it is because the information about the brightness of the pixel which takes extremal value becomes indistinct gradually with those positional information. Drawing 1 (a) and drawing 1 (b) show the

result of having applied the equalization filter to the image of a face. While the brightness of the pixel which takes extremal value as shown in this drawing fades gradually by equalization, the location is also shifted under the effect of equalization. Consequently, neither the brightness of an eye (the minimum point of brightness) nor the information on a location can become ambiguous on such coarse resolution level, and cannot calculate right matching in this resolution. Therefore, though it is for global matching preparing coarse resolution level, matching obtained here does not correspond to the true description (an eye, i.e., the minimum point) of an image correctly. Even if an eye appears vividly on more minute resolution level, recovery does not attach any longer the error mixed when taking global matching. By adding smoothing processing to an input image, it is also already pointed out that the stereophonic information of a texture field falls.

[0031] On the other hand, there is a "screen (sieve)" operator of a single dimension as a nonlinear filter which began to be used in the field of geomorphology recently. This operator adds smoothing processing to an image by choosing the minimal value in the single dimension window of predetermined magnitude (or maximal value), saving the causal relation of a scale and space. Although the image obtained as a result is the same magnitude as the original image, since the component of a small wave is removed, it becomes simpler. Although this operator can be classified into the "multiplex resolution filter" in large semantics in that the information on an image is dropped, an image is not necessarily (it being got blocked and not being a multiplex resolution filter in narrow semantics) hierarchized, changing the resolution of an image like wavelet in fact, and it cannot use for detection of correspondence between images.

[0032] If beyond [the technical problem which a premise technique tends to solve] is summarized, the following technical problems will be accepted.

1. The image-processing approach of grasping the description of an image by comparatively easy processing correctly was scarce. There were few effective proposals about the image-processing approach that the description can be extracted maintaining the information about an especially characteristic point, for example, a pixel value, and a location.
2. When corresponding points were detected automatically based on the description of an image, there was a fault, like generally processing is complicated or noise resistance is low. Moreover, it was difficult to take matching which needs to prepare various constraint on the occasion of processing, and fulfills the global description and the local description to coincidence.
3. In order to recognize the global structure or the global description of an image, even if it introduced the multiplex resolution filter, when the filter was a linearity filter, the brightness information and positional information of a pixel became ambiguous. Consequently, grasp of corresponding points tended to become incorrectness. Since

the single dimension screen operator which is a nonlinear filter did not hierarchize an image, it was not applicable to detection of the corresponding points between images.

4. When it was going to grasp corresponding points correctly these results, it depended on assignment by the help after all, and also there were no effective means.

[0033] A premise technique is made for the purpose of solution of these technical problems, and offers the technique which enables exact grasp of the description of an image in the field of an image processing.

[0034] [-- a premise technique -- The means for solving a technical problem] -- the mode which has a premise technique for this purpose proposes the image filter of new multiplex resolution. This multiplex resolution filter extracts the singular point from an image. Therefore, it is also called a singular point filter. The singular point means a point with the image top description. About the maximum point that a pixel value (a pixel value points out numeric values of the arbitration about an image or a pixel, such as a color number number and a brightness value) becomes max in a certain field as an example, the minimum point which becomes min, and a certain direction, although it is max, there is the saddle point which becomes min about another direction. The singular point may be a concept on topology. It may correct, in addition you may have what kind of description. It is not an essential problem for a premise technique of what kind of property a point is considered to be the singular point.

[0035] In this mode, the image processing which used the multiplex resolution filter is performed. First, in a detection process, to the first image, two-dimensional retrieval is performed and the singular point is detected. Next in a generation process, the detected singular point is extracted, and the second image with resolution lower than the first image is generated. The singular point which the first image has is taken over to the second image. Since resolution is lower than the first image, the second image is suitable for grasp of the global description of an image.

[0036] Another mode of a premise technique is related with the image matching approach of having used the singular point filter. In this mode, matching between a starting point image and a terminal point image is taken. A starting point image and a terminal point image are the names conferred for convenience because of distinction of two images, and there is no essential difference.

[0037] In this mode, a series of starting point hierarchy images with which a singular point filter is given to a starting point image, and resolution differs at the first process are generated. At the second process, a series of terminal point hierarchy images with which a singular point filter is given to a terminal point image, and resolution differs are generated. A starting point hierarchy image and a terminal point hierarchy image mean the image group which hierarchizes a starting point image and a terminal point image, respectively, and is obtained, and it consists of an image of at least two sheets, respectively. Next in the third process, matching of a starting point hierarchy image and a terminal point hierarchy image is calculated in the hierarchy of resolution level.

Since according to this mode the description of the image relevant to the singular point is extracted by the multiplex resolution filter and/or it is clarified, matching becomes easy. Especially the constraint for matching is not needed.

[0038] Still more nearly another mode of a premise technique is also related with matching of a starting point image and a terminal point image. A valuation plan is beforehand formed about each of two or more matching evaluation criteria, those valuation plans are unified, and this mode defines a comprehensive valuation plan, and is searched for the optimal matching paying attention to near the extremal value of that comprehensive valuation plan. A comprehensive valuation plan may be defined as total of those valuation plans, after [of a valuation plan] hanging a multiplier parameter on one at least, it may detect the condition that a comprehensive valuation plan or one of valuation plans take extremal value mostly in that case, and may determine said parameter. It is because the error may be included "near extremal value" or to have presupposed [some] "Extremal value is taken mostly." Some errors do not very pose a problem at a premise technique.

[0039] In order for the extremal value itself to be dependent on said parameter, room to determine the parameter considered to be the optimal based on the behavior of extremal value, i.e., the situation of change of extremal value, arises. This mode uses that fact. According to this mode, the way which automates the decision of the difficult parameter of adjustment originally is cultivated.

[0040] The component engineering of a premise technique explains in full detail by [1] to the [gestalt of operation of premise technique] beginning, and [2] explains procedure concretely. Furthermore, [3] reports the result of an experiment.

[0041] [1] Introduce the new multiplex resolution filter called the detail [1.1] introduction singular point filter of the component engineering, and calculate matching between images correctly. The preliminary knowledge about an object is unnecessary entirely. Count of matching between images is calculated in each resolution, while progressing the hierarchy of resolution. The hierarchy of resolution is followed from coarse level in order to minute level in that case. A parameter required for count is completely set automatically by the dynamic count similar to human being's vision system. It is not necessary to specify the corresponding points between images with a help.

[0042] This premise technique is applicable to generation of the smooth dynamic image from completely automatic morphing, an object recognition, stereophotogrammetry, volume rendering, and few frames etc., for example. When using for morphing, the given image can be transformed automatically. When using for volume rendering, the in-between image between cross sections can be reconstructed correctly. It is the same, even when the configuration of a long distance and a cross section is large and the distance between cross sections changes.

[0043] [1.2] The multiplex resolution singular point filter concerning the hierarchy

premise technique of a singular point filter can save the brightness and location of each singular point which are moreover included in an image, dropping the resolution of an image. Width of face of an image is set to N, and height is set to M here. Below, since it is easy, it is assumed that it is $N=M=2n$ (n is the natural number). Moreover, $[0$ and Section N] $**R$ is described to be I. The pixel of the image in (i, j) is described to be $p(i, j)$ ($i, j \leq I$).

[0044] The hierarchy of multiplex resolution is introduced here. The hierarchized image group is generated by the multiplex resolution filter. A multiplex resolution filter generates another image with resolution lower than the image of even if it performs two-dimensional retrieval to the image of a basis, it detects the singular point and it extracts the detected singular point. Size of each image in the m -th level is set to $2m \times 2m$ ($0 \leq m \leq n$) here. A singular point filter builds recursively four kinds of new hierarchy images as follows towards falling from n .

[0045]

[Equation 1]

$$\begin{aligned} p_{(i,j)}^{(m,0)} &= \min(\min(p_{(2i,2j)}^{(m+1,0)}, p_{(2i,2j+1)}^{(m+1,0)}), \min(p_{(2i+1,2j)}^{(m+1,0)}, p_{(2i+1,2j+1)}^{(m+1,0)})) \\ p_{(i,j)}^{(m,1)} &= \max(\min(p_{(2i,2j)}^{(m+1,1)}, p_{(2i,2j+1)}^{(m+1,1)}), \min(p_{(2i+1,2j)}^{(m+1,1)}, p_{(2i+1,2j+1)}^{(m+1,1)})) \\ p_{(i,j)}^{(m,2)} &= \min(\max(p_{(2i,2j)}^{(m+1,2)}, p_{(2i,2j+1)}^{(m+1,2)}), \max(p_{(2i+1,2j)}^{(m+1,2)}, p_{(2i+1,2j+1)}^{(m+1,2)})) \\ p_{(i,j)}^{(m,3)} &= \max(\max(p_{(2i,2j)}^{(m+1,3)}, p_{(2i,2j+1)}^{(m+1,3)}), \max(p_{(2i+1,2j)}^{(m+1,3)}, p_{(2i+1,2j+1)}^{(m+1,3)})) \end{aligned}$$

(式 1)

However, it is here and is [Equation 2].

$$p_{(i,j)}^{(n,0)} = p_{(i,j)}^{(n,1)} = p_{(i,j)}^{(n,2)} = p_{(i,j)}^{(n,3)} = p_{(i,j)} \quad (\text{式 2})$$

It carries out. These four images are henceforth called a subimage (subimage). If $\min x \leq t \leq x+1$ and $\max x \leq t \leq x+1$ are described to be alpha and beta, respectively, a subimage can be described as follows, respectively.

[0046]

$$P(m, 0) = \alpha(x) \alpha(y) p(1 \text{ m} + 0) \quad P(m, 1) = \alpha(x) \beta(y) p(1 \text{ m} + 1) \\ P(m, 2) = \beta(x) \alpha(y) p(1 \text{ m} + 2)$$

$P(m, 3) = \beta(x) \beta(y) p(1 \text{ m} + 3)$, i.e., these, is considered to be a thing like the tensor product of alpha and beta. The subimage supports the singular point, respectively. A singular point filter detects the singular point for every block which consists of 2×2 pixels about the image of a basis so that clearly from these formulas.

It searches for a point with the maximum pixel value or the minimum pixel value about two directions, i.e., the length, and the side of each block in that case. As a pixel value, although brightness is adopted with a premise technique, various numeric values about an image are employable. While the pixel from which the pixel which serves as the maximum pixel value about both two directions serves as the minimum pixel value about both the maximum point and two directions serves as the maximum pixel value about one side of the minimum point and two directions, the pixel which serves as the minimum pixel value about another side is detected as the saddle point.

[0047] A singular point filter drops the resolution of an image by representing the image (here 4 pixels) of the block with the image (here 1 pixel) of the singular point detected inside each block. If it carries out from a theoretical viewpoint of the singular point, $\alpha(x)$ $\alpha(y)$ saves the minimum point, $\beta(x)$ $\beta(y)$ saves the maximum point and $\alpha(x)$ $\beta(y)$ and $\beta(x)$ $\alpha(y)$ saves the saddle point.

[0048] First, singular point filtering is separately performed to the starting point (source) image and terminal point (destination) image which should take matching, and a series of image groups, i.e., starting point hierarchy image, and terminal point hierarchy images are generated, respectively. A starting point hierarchy image and four kinds of terminal point hierarchy images are generated at a time corresponding to the class of singular point, respectively.

[0049] Then, matching of a starting point hierarchy image and a terminal point hierarchy image can be taken in a series of resolution level, and it is *****. Matching of the minimum point is first taken using $p(m, 0)$. Next, based on the result, matching of the saddle point is taken using $p(m, 1)$, and matching of other saddle points is taken using $p(m, 2)$. And finally $p(m, 3)$ is used and matching of the maximum point is taken.

[0050] Drawing 1 (c) and drawing 1 (d) show the subimage $p(5\ 0)$ of drawing 1 (a) and drawing 1 (b), respectively. Similarly, as for drawing 1 (e) and drawing 1 (f), in $p(5\ 1)$, drawing 1 (g), and drawing 1 (h), $p(5\ 2)$, drawing 1 (i), and drawing 1 (j) show $p(5\ 3)$, respectively. According to the subimage, matching of the description part of an image becomes easy as these drawings show. An eye becomes clear by $p(5\ 0)$ first. It is because an eye is the minimum point of brightness in a face. According to $p(5\ 1)$, opening becomes clear. Opening is because brightness is low in a longitudinal direction. According to $p(5\ 2)$, the vertical line of the both sides of a neck becomes clear. The brightest point of a lug or a cheek becomes clear by $p(5\ 3)$ at the last. It is because these are the maximum points of brightness.

[0051] Since the description of an image can be extracted according to the singular point filter, the photographic subject reflected in a camera is discriminable by comparing the description of some objects beforehand recorded as the description of the image photoed with the camera.

[0052] [1.3] Write the pixels of the location (i, j) of the count starting point image of

the map between images to be $p(n)$, and (i, j) , and, similarly $q(n)$, and (k, l) describe the pixel of the location (k, l) of a terminal point image. It considers as i, j, k , and $l \in I$. The energy (after-mentioned) of the map between images is defined. This energy is decided by the difference of the brightness of the pixel of a starting point image, and the brightness of the pixel to which a terminal point image corresponds, and smoothness of a map. Map $f(m, 0): p(m, 0) \rightarrow q(m, 0)$ between $p(m, 0)$ which has the minimum energy in the beginning, and $q(m, 0)$ is calculated. Based on $f(m, 0)$, the map $f(m, 1)$ between $p(m, 1)$ with the threshold energy and $q(m, 1)$ is calculated. This procedure continues until count of the map $f(m, 3)$ between $p(m, 3)$ and $q(m, 3)$ is completed. Each map $f(m, i)$ ($i = 0, 1$ and $2, \dots$) will be called a submap. For the convenience of count of $f(m, i)$, the sequence of i can be rearranged like a degree type. A reason to be rearranged is mentioned later.

[0053]

[Equation 3]

$$f^{(m,i)} : p^{(m,\sigma(i))} \rightarrow q^{(m,\sigma(i))} \quad (\text{式 3})$$

It is $\sigma(i) \in \{0, 1, 2, 3\}$ here.

[0054] [1.3.1] When expressing matching between a bijection starting point image and a terminal point image by map, the map should fulfill bijection conditions among both images. It is because there are no superiority or inferiority on a concept in both images and a mutual pixel should be connected by surjection and the monomorphism. However, the map which should be built here unlike the usual case is the digital version of bijection. A pixel is specified by the lattice point with a premise technique.

[0055] The map to a terminal point secondary image (subimage prepared about the terminal point image) from a starting point secondary image (subimage prepared about the starting point image) is expressed by $f(m, s): I/2 \times n-m \times I/2 \times n-m \rightarrow I/2 \times n-m \times I/2 \times n-m$ ($s = 0, 1, \dots$). Here, $f(m, s)(i, j) = (k, l)$ means that $p(m, s)(i, j)$ of a starting point image is mapped by $q(m, s)(k, l)$ of a terminal point image. Since it is easy, when $f(i, j) = (k, l)$ is realized, Pixel $q(k, l)$ is described to be $qf(i, j)$.

[0056] Like the pixel (lattice point) treated with a premise technique, when data are discrete, the definition of bijection is important. Here, a definition is given as follows (all of i, i', j, j' , and k and l consider as an integer). Each square field, [Equation 4] which are first written by R in the flat surface of a starting point image

$$p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)} \quad (\text{式 4})$$

***** ($i = 0, \dots, 2m-1, j = 0, \dots, 2m-1$). The direction of each side (edge) of R is defined as follows here.

[0057]

[Equation 5]

$$\overrightarrow{p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)}}, \overrightarrow{p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)}}, \overrightarrow{p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)}} \text{ and } \overrightarrow{p_{(i,j+1)}^{(m,s)} p_{(i,j)}^{(m,s)}} \quad (\text{式 5})$$

This square must be mapped by Map f by the quadrilateral in a terminal point image plane. The quadrilateral shown by f (m, s) (R), [Equation 6]

$$q_f^{(m,s)}(i,j) q_f^{(m,s)}(i+1,j) q_f^{(m,s)}(i+1,j+1) q_f^{(m,s)}(i,j+1) \quad (\text{式 6})$$

It is necessary to fulfill ** and the following bijection conditions.

[0058] 1. The edge of quadrilateral f (m, s) (R) does not cross mutually.

2. The direction of the edge of f (m, s) (R) is equal to them of R (in the case of drawing 2 clockwise rotation).

3. Allow a contraction map (RITORA cushion: retractions) as relaxation conditions.

[0059] Unless some relaxation conditions are established, the map which fulfills bijection conditions completely is because there is only a unit map. Here, as for 0 (m, s) (R), i.e., f, the die length of one edge of f (m, s) (R) may become a triangle. However, don't become a graphic form with which area is set to 0, i.e., one point, and one segment. Although drawing 2 (A) and drawing 2 (D) fulfill bijection conditions when drawing 2 (R) is the quadrilateral of a basis, drawing 2 (B), drawing 2 (C), and drawing 2 (E) do not fill.

[0060] In an actual implementation, the following conditions may be imposed further that it should guarantee easily that a map is surjection. That is, in each pixel on the boundary of a starting point image, the pixel which occupies the same location in a terminal point image ****. That is, it is f(i, j) = (i, j) (however, on four lines of i= 0, i=2m-1, j= 0, and j=2m-1). This condition is also called "addition conditions" below.

[0061] [1.3.2] Define the energy of the cost map f about the brightness of the energy [1.3.2.1] pixel of a map. It is the purpose to look for the map to which energy becomes min. Energy is mainly decided by the difference of the brightness of the pixel of a starting point image, and the brightness of the pixel of the terminal point image corresponding to it. That is, the energy C (m, s) in the point (i, j) of Map f (m, s) (i, j) becomes settled by the degree type.

[0062]

[Equation 7]

$$C_{(i,j)}^{(m,s)} = |V(p_{(i,j)}^{(m,s)}) - V(q_f^{(m,s)}(i,j))|^2 \quad (\text{式 7})$$

Here, V (p (m, s) (i, j)) and V (q(m, s) f (i, j)) are the brightness of Pixel p (m, s) (i, j) and

$q(m, s) f(i, j)$, respectively. The total energy C of $f(m, s)$ is one valuation plan by which matching is evaluated, and the sum total of $C(m, s)(i, j)$ shown below can define it.

[0063]

[Equation 8]

$$C_f^{(m,s)} = \sum_{i=0}^{i=2^m-1} \sum_{j=0}^{j=2^m-1} C_{(i,j)}^{(m,s)} \quad (\text{式 } 8)$$

[1.3.2.2] In order to obtain the cost smooth map about the location of the pixel for a smooth map, introduce another energy D_f about a map. This energy is decided by the location of $p(m, s)(i, j)$ and $q(m, s) f(i, j)$ regardless of the brightness of a pixel ($i=0, \dots, 2^m-1, j=0, \dots, 2^m-1$). The energy $D(m, s)$ of the map $f(m, s)$ in a point (i, j) is defined by the degree type.

[0064]

[Equation 9]

$$D_{(i,j)}^{(m,s)} = \eta E_{0(i,j)}^{(m,s)} + E_{1(i,j)}^{(m,s)} \quad (\text{式 } 9)$$

However, the multiplier parameters η are the zero or more real numbers, and are [Equation 10].

$$E_{0(i,j)}^{(m,s)} = \|(i, j) - f^{(m,s)}(i, j)\|^2 \quad (\text{式 } 10)$$

[Equation 11]

$$E_{1(i,j)}^{(m,s)} = \sum_{i'=i-1}^i \sum_{j'=j-1}^j \|(f^{(m,s)}(i, j) - (i, j)) - (f^{(m,s)}(i', j') - (i', j'))\|^2 / 4 \quad (\text{式 } 11)$$

It carries out. It is here and is [Equation 12].

$$\|(x, y)\| = \sqrt{x^2 + y^2} \quad (\text{式 } 12)$$

It comes out, and it is and $f(i', j')$ decides on 0 to $i' < 0$ and $j' < 0$. E_0 and (i, j) it is decided by distance of $f(i, j)$. It prevents E_0 to the pixel which the pixel left too much. However, E_0 is replaced with behind another energy function. E_1 guarantees the smoothness of a map. E_1 expresses the distance between the variation rates and the variation rates of an adjacent point of $p(i, j)$. Based on the above consideration, the energy D_f which is another valuation plan by which matching is evaluated becomes settled in a degree type.

[0065]

[Equation 13]

$$D_f^{(m,s)} = \sum_{i=0}^{2^m-1} \sum_{j=0}^{2^m-1} D_{(i,j)}^{(m,s)} \quad (\text{式 1 3})$$

[1.3.2.3] The comprehensive valuation plan concerning integration of the gross energy of a gross energy map of a map, i.e., two or more valuation plans, is defined by $\lambda C(m, s) + D(m, s) f$. The multiplier parameters λ are the zero or more real numbers here. The purpose is detecting the condition a comprehensive valuation plan's taking extremal value, i.e., finding out the map which gives the threshold energy shown by the degree type.

[0066]

[Equation 14]

$$\min_f \lambda C_f^{(m,s)} + D_f^{(m,s)} \quad (\text{式 1 4})$$

In the case of $\lambda = 0$ and $\eta = 0$, it should be cautious of a map turning into a unit map (that is, it becomes $f(m, s)(i, j) = (i, j)$ to all $i = 0, \dots, 2^m-1$ and $j = 0, \dots, 2^m-1$). Since this premise technique estimates the case of $\lambda = 0$ and $\eta = 0$ first like the after-mentioned, a map can be gradually transformed from a unit map. If the location of λ of a comprehensive valuation plan was changed temporarily and it was defined as $C(m, s) + \lambda D(m, s) f$, the pixels which in the case of $\lambda = 0$ and $\eta = 0$ a comprehensive valuation plan is set only to $C(m, s) f$, and originally do not have relation in any way are matched only by saying [that brightness is near], and they become what has a meaningless map. Even if it transforms the map based on such a meaningless map, semantics is not made at all. For this reason, it is at the initiation time of evaluation of a unit map, and is considered in how to give a multiplier parameter so that it may be chosen as the best map.

[0067] The difference and smoothness of brightness of a pixel are taken into consideration like [an optical flow] this premise technique. However, an optical flow cannot be used for conversion of an image. It is for taking only a local motion of an object into consideration. Global correspondence relation is detectable by using the singular point filter concerning a premise technique.

[0068] [1.3.3] Give the decision threshold energy of the map by installation of multiplex resolution, and ask for the map f_{\min} which satisfies bijection conditions using the hierarchy of multiplex resolution. In each resolution level, the map between a starting point secondary image and a terminal point secondary image is calculated. It starts from the most significant (coarsest level) of the hierarchy of resolution, and the map of each resolution level is determined, taking the map of other level into consideration. The number of the candidates of the map in each level is restricted by

using the map of coarser level more highly that is,. More specifically on the occasion of the decision of the map in a certain level, the map called for in level coarser [one] than it is imposed as a kind of constraint.

[0069] First, [Equation 15]

$$(i', j') = ([\frac{i}{2}], [\frac{j}{2}]) \quad (\text{式 1 5})$$

***** -- it comes suddenly and p (m-1, s) (i', j') and q (m-1, s) (i', j') will be called parent of p (m, s) (i, j) and q (m, s) (i, j), respectively. [x] is the maximum integer which does not exceed x. Moreover, p (m, s) (i, j) and q (m, s) (i, j) are called child of p (m-1, s) (i', j') and q (m-1, s) (i', j'), respectively. Function parent (i, j) is defined by the degree type.

[0070]

[Equation 16]

$$parent(i, j) = ([\frac{i}{2}], [\frac{j}{2}]) \quad (\text{式 1 6})$$

It is determined by finding what the map f (m, s) between p (m, s) (i, j) and q (m, s) (k, l) performed energy count, and became min. By using f (m-1, s) (m= 1, 2, --, n), the value of f(m, s) (i, j) = (k, l) is determined as follows. First, q (m, s) (k, l) imposes the conditions that it must be in the interior of the following quadrilateral, and narrows down what has high actuality among the maps which fulfill bijection conditions.

[0071]

[Equation 17]

$$q_{g^{(m,s)}(i-1,j-1)}^{(m,s)} q_{g^{(m,s)}(i-1,j+1)}^{(m,s)} q_{g^{(m,s)}(i+1,j+1)}^{(m,s)} q_{g^{(m,s)}(i+1,j-1)}^{(m,s)} \quad (\text{式 1 7})$$

However, it is here and is [Equation 18].

$$g^{(m,s)}(i, j) = f^{(m-1,s)}(parent(i, j)) + f^{(m-1,s)}(parent((i, j) + (1, 1))) \quad (\text{式 1 8})$$

It comes out. In this way, the defined quadrilateral is made to call it the inheritance (inherited) quadrilateral of Following p (m, s) (i, j). It asks for the pixel which makes energy min in the interior of an inheritance quadrilateral.

[0072] Drawing 3 shows the above procedure. said -- drawing -- setting -- the starting point -- an image -- A -- B -- C -- D -- a pixel -- the -- m - one -- level -- setting -- respectively -- a terminal point -- an image -- A -- ' -- B -- ' -- C -- ' -- D -- ' -- ****(ing) -- having . Pixel p (m, s) (i, j) must be ****(ed) to pixel q(m, s)

$f(m)$ which exists in the interior of inheritance quadrilateral $A'B'C'D'$, and (i, j) . The mediation to the map of the m -th level from a map of the $m-1$ st level is made by the above consideration.

[0073] The energy E_0 defined previously is transposed to a degree type, in order to calculate the submap f in the m -th level $(m, 0)$.

[0074]

[Equation 19]

$$E_{0(i,j)} = ||f^{(m,0)}(i,j) - g^{(m)}(i,j)||^2 \quad (\text{式 } 19)$$

Moreover, a degree type is used in order to calculate the submap $f(m, s)$.

[0075]

[Equation 20]

$$E_{0(i,j)} = ||f^{(m,s)}(i,j) - f^{(m,s-1)}(i,j)||^2 \quad (1 \leq i) \quad (\text{式 } 20)$$

In this way, the map which maintains the energy of all submaps at a low value is obtained. It is related within the same level by the formula 20 so that the similarity of submaps may become [the submap corresponding to the different singular point] high. The formula 19 shows distance with the location of the point that (i, j) at the time of considering it a part of $f(m, s)(i, j)$ and pixel of the $m-1$ st level should be projected. [0076] The following measures are taken when the pixel which fulfills bijection conditions does not exist in the interior of inheritance quadrilateral $A'B'C'D'$ temporarily. First, the pixel whose distance from the boundary line of $A'B'C'D'$ is L (it is $L=1$ in the beginning) is investigated. Among those, if that from which energy becomes min fulfills bijection conditions, this will be chosen as a value of $f(m, s)(i, j)$. L is enlarged until such a point is discovered or L reaches $L(m)$ max of the upper limit. $L(m)$ max is immobilization to each level m . When such a point is not discovered at all, a map to which the 3rd condition of bijection is disregarded temporarily and the area of the quadrilateral of a conversion place becomes zero is also accepted, and $f(m, s)(i, j)$ is determined. When the point of still fulfilling conditions is not found, the 1st and 2nd conditions of bijection are removed next.

[0077] Avoiding that a map is influenced by the details of an image, the approximation using multiplex resolution is indispensable in order to determine the global correspondence relation between images. If approximation by multiplex resolution is not used, it is impossible to find out the correspondence relation between pixels with a far distance. In that case, the size of an image must be limited to a very small thing, and can treat only the small image of change. Furthermore, correspondence relation between such pixels is made hard to find it, in order to usually require smoothness of a map. The energy of the map to a pixel with distance from a pixel is because it is high.

According to the approximation using multiplex resolution, the suitable correspondence relation between such pixels can be found out. Those distance is because it is small in the upper level (coarse level) of the hierarchy of resolution.

[0078] [1.4] There is difficulty [the main faults of the matching technique of automatic decision existing of the optimal PAREMETA value] of one PAREMETA adjustment. When the most, it is very difficult for adjustment of a parameter to be performed by the help activity and to choose the optimal value. According to the approach concerning a premise technique, an automatic decision of the optimal parameter value can be made completely.

[0079] The system concerning a premise technique contains two PAREMETA, and lambda and eta. Speaking directly, lambda's being the weight of the difference of the brightness of a pixel and eta's showing the rigidity of a map. Initial value is 0, it fixes to eta= 0 first and the value of these parameters makes lambda increase from 0 gradually. When making the value of a comprehensive valuation plan (formula 14) into min moreover, enlarging the value of lambda, generally the value of $C(m, s)$ f about each ***** becomes small. This means that two images must match more fundamentally. However, if lambda exceeds an optimum value, the following phenomena will occur.

[0080] 1. The pixels which should not correspond essentially are accidentally matched only by saying [that brightness is near].

2. Consequently, the correspondence relation of pixels becomes amusing and a map begins to collapse.

[0081] 3. Consequently, in a formula 14, $D(m, s)$ f tends to increase rapidly.

4. Consequently, since the value of a formula 14 tends to increase rapidly, $f(m, s)$ changes so that the rapid increment in $D(m, s)$ f may be controlled, and, as a result, $C(m, s)$ f increases. Therefore, maintaining the condition of saying that the minimum value is taken making lambda increase, $C(m, s)$ f detects the threshold changed to an increment from reduction, and a formula 14 makes the lambda the optimum value in eta= 0. Next, eta is increased little by little, the behavior of $C(m, s)$ f is inspected, and an automatic decision of eta is made by the below-mentioned approach. lambda is also decided corresponding to the eta.

[0082] This approach resembles actuation of the focal device human being's vision system. In human being's vision system, matching of the image of right-and-left both eyes is taken, moving one eye. The eye is fixed when an object can recognize clearly.

[0083] [1.4.1] It is increased from 0 by the dynamic determination lambda of lambda by predetermined unit width of face, and a submap is evaluated whenever the value of lambda changes. Gross energy is defined by $\lambda C(m, s) f + D(m, s) f$ like a formula 14. $D(m, s) f$ of a formula 9 expresses smoothness, and, in a unit map, becomes min theoretically, and $E0$ and $E1$ increase it, so that a map is distorted. Since $E1$ is an integer, the minimum unit width of face of $D(m, s) f$ is 1. For this reason, if change (decrement) of present $\lambda C(m, s) (i, j)$ is not one or more, gross energy cannot be

reduced by changing a map. Because, since $D(m, s)$ increases one or more with change of a map, gross energy is for not decreasing unless $\lambda C(m, s)(i, j)$ decreases or more by one.

[0084] With the basis of this condition, and the increment in λ , when normal, it is shown that $C(m, s)(i, j)$ decreases. The histogram of $C(m, s)(i, j)$ is described to be $h(l)$. $h(l)$ is a number which is the pixel whose energy $C(m, s)(i, j)$ is l . Since $\lambda l \geq 1$ is realized, the case of $l = 1/\lambda$ is considered. It is [Equation 21] when λ carries out minute amount change from λ_1 to λ_2 .

$$A = \sum_{l=\lceil \frac{1}{\lambda_2} \rceil}^{\lfloor \frac{1}{\lambda_1} \rfloor} h(l) \simeq \int_{\frac{1}{\lambda_2}}^{\frac{1}{\lambda_1}} h(l) dl = - \int_{\lambda_2}^{\lambda_1} h(l) \frac{1}{\lambda^{3/2}} d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{h(l)}{\lambda^{3/2}} d\lambda$$

(式 2 1)

A pixels come out of and shown are [Equation 22].

$$C_f^{(m,s)} - l^2 = C_f^{(m,s)} - \frac{1}{\lambda}$$

(式 2 2)

It changes [rather than] to a stable condition with ** energy. Here, temporarily, if all of the energy of these pixels become zero, it approximates. For this formula, the value of $C(m, s)$ is [Equation 23].

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{1}{\lambda}$$

(式 2 3)

Carrying out **** change is shown, consequently it is [Equation 24].

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{h(l)}{\lambda^{5/2}}$$

(式 2 4)

It ***** $h(l) > 0$. Since it is 0, $C(m, s)$ usually decreases. However, when λ tends to exceed an optimum value, an above-mentioned phenomenon, i.e., the increment in $C(m, s)$, occurs. The optimum value of λ is determined by detecting this phenomenon.

[0085] In addition, it is [Equation 25] when making H ($h > 0$) and k into a constant.

$$h(l) = H l^k = \frac{H}{\lambda^{k/2}}$$

(式 2 5)

It will be [Equation 26] if it assumes.

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{H}{\lambda^{5/2+k/2}}$$

(式 2 6)

*****. It is [Equation 27] if it is $k! = -3$ at this time.

$$C_f^{(m,s)} = C + \frac{H}{(3/2 + k/2)\lambda^{3/2+k/2}}$$

It becomes. This is the general formula of $C(m, s)$ (C is a constant).

[0086] In case the optimum value of λ is detected, insurance may be seen further and the number of the pixels which break bijection conditions may be inspected. In case the map of each pixel is determined here, the probability to break bijection conditions is assumed to be p_0 . In this case, [Equation 28]

$$\frac{\partial A}{\partial \lambda} = \frac{h(l)}{\lambda^{3/2}} \quad (\text{式 28})$$

Since it is ***** (ing), the number of the pixels which break bijection conditions increases at the rate of a degree type.

[0087]

[Equation 29]

$$B_0 = \frac{h(l)p_0}{\lambda^{3/2}} \quad (\text{式 29})$$

Therefore, [Equation 30]

$$\frac{B_0 \lambda^{3/2}}{p_0 h(l)} = 1 \quad (\text{式 30})$$

It is ***** [Equation 31] when assuming $h(l) = Hl$ temporarily

$$B_0 \lambda^{3/2+k/2} = p_0 H \quad (\text{式 31})$$

It becomes ***** However, if λ exceeds an optimum value, the upper value will increase quickly. This phenomenon can be detected, it can inspect whether a $B_0 \lambda^{3/2+k/2}/2m$ value exceeds outlying-observation $B_0 \text{thres}$, and the optimum value of λ can be determined. The rate of increase B_1 of a pixel which breaks the 3rd condition of bijection is checked by similarly inspecting whether a $B_1 \lambda^{3/2+k/2}/2m$ value exceeds outlying-observation $B_1 \text{thres}$. The reason for introducing factor $2m$ is mentioned later. This system is not sensitive to these two thresholds. By observation of energy $C(m, s)$, these thresholds can be used in order to detect too much distortion of the map which it failed to detect.

[0088] In addition, in the experiment, when calculating the submap $f(m, s)$, supposing

lambda exceeded 0.1, count of f (m, s) was stopped and shifted to count of f (m, s+1). At the time of lambda > 0.1, it is because [in the brightness 255 level of a pixel] the difference in "3" influenced count of a submap only, and it is because it was difficult to obtain a right result at the time of lambda > 0.1.

[0089] [1.4.2] Histogram h (l)

It does not depend for inspection of C(m, s) f on histogram h (l). Effect can be received in h (l) in the case of inspection of bijection and its 3rd condition. When it actually (lambda, C(m, s) f) plots, k is usually in the one neighborhood. In the experiment, B0lambda2 and B1lambda2 were inspected using k= 1. With [the true value of k] one [less than], B0lambda2 and B1lambda2 do not become a constant, but they increase gradually according to factor lambda (1-k) / 2. If h (l) is a constant, factors are lambda1/2, for example. However, such a difference is absorbable by setting up threshold B0thres correctly.

[0090] A core assumes a starting point image to be (x0, y0), and the circular object of a radius r like a degree type here.

[Equation 32]

$$p(i,j) = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_0)^2 + (j-y_0)^2}) & (\sqrt{(i-x_0)^2 + (j-y_0)^2} \leq r) \\ 0 & (otherwise) \end{cases} \quad (式 3 2)$$

On the other hand, a terminal point image presupposes that a core (x1, y1) and a radius are the objects of r like a degree type.

[Equation 33]

$$q(i,j) = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_1)^2 + (j-y_1)^2}) & (\sqrt{(i-x_1)^2 + (j-y_1)^2} \leq r) \\ 0 & (otherwise) \end{cases} \quad (式 3 3)$$

c (x) presupposes that it is the form of c(x) = xk here. A core (x0, y0) and (x1, y1) when sufficiently far, histogram h (l) becomes the form of a degree type.

[Equation 34]

$$h(l) \propto r l^k \quad (k \neq 0) \quad (式 3 4)$$

An image shows an object with the clear boundary line embedded for the background at the time of k= 1. This object becomes bright as it is dark in a core and it goes to a perimeter. An image expresses an object with an ambiguous boundary line at the time of k=-1. This object becomes dark as it is the brightest in a core and it goes to a perimeter. Even if it thinks that there is a general object in the middle of these two types of object, generality is not very lost. Therefore, k can cover the case where it is the most as -1 <= k <= 1, and it is secured that a formula 27 is generally a decreasing function.

[0091] In addition, as shown in a formula 34, it should be cautious of r being influenced by the resolution of an image, i.e., r being proportional to $2m$. For this reason, factor $2m$ was introduced in [1.4.1].

[0092] [1.4.3] Make an automatic decision of the dynamic determination parameter η of η by the same approach. It is first referred to as $\eta = 0$, and final map $f(n)$ and Energy $C(n)$ f in the finest resolution are calculated. Continue, only a certain value $\Delta\eta$ makes η increase, and last map $f(n)$ and Energy $C(n)$ f in the again finest resolution are recalculated. This process is continued until an optimum value can be found. η shows the rigidity of a map. It is because it is the weight of a degree type.

[0093]

[Equation 35]

$$E_{0(i,j)}^{(m,s)} = ||f^{(m,s)}(i,j) - f^{(m,s-1)}(i,j)||^2$$

(式 3 5)

When η is 0, $D(n)$ f will be determined regardless of the last submap, the present submap will deform elastically, and it will be distorted too much. On the other hand, as for $D(n)$ f , η is decided by the last submap nearly completely at the very big time of a value. At this time, a submap has very high rigidity and a pixel is projected on the same location. Consequently, a map turns into a unit map. When the value of η increases gradually from 0, $C(n)$ f decreases gradually like the after-mentioned. However, energy begins to increase as it is shown in drawing 4, when the value of η exceeds an optimum value. The X-axis of this drawing is η and a Y-axis is Cf .

[0094] The value of optimal η which makes $C(n)$ f min by this approach can be acquired. However, as a result of various elements' influencing count compared with the case of λ , $C(n)$ f changes with fluctuation small. In η , it is because all submaps are recalculated but as a submap is recalculated once, whenever an input carries out minute amount change of the case of λ . For this reason, it cannot judge immediately whether the value of obtained $C(n)$ f is min. If the candidate of the minimum value is found, it is necessary to look for the true minimum value by setting up the still finer section.

[0095] [1.5] In case the correspondence relation between super sampling pixels is determined, in order to increase a degree of freedom, the range of $f(m, s)$ is extensible to $R \times R$ (R is the set of the real number). In this case, the brightness of the pixel of a terminal point image is interpolated and they are a nonintegral point and [Equation 36].

$$V(q_{f^{(m,s)}(i,j)}^{(m,s)}) \quad (式 3 6)$$

$f(m, s)$ with the brightness which can be boiled and set is offered. That is, a super

sampling is performed. By experiment, for $f(m, s)$, it is allowed and it is [Equation 37] to take an integer and a half-integral value.

$$V(q_{(i,j)+(0.5,0.5)}^{(m,s)}) \quad (\text{式 3 7})$$

**, [Equation 38]

$$(V(q_{(i,j)}^{(m,s)}) + V(q_{(i,j)+(1,1)}^{(m,s)}))/2 \quad (\text{式 3 8})$$

Be alike was given.

[0096] [1.6] When the normalization starting point image and terminal point image of brightness of each image contain an extremely different object, if the brightness of the original pixel remains as it is to count of a map, it is hard to use for it. [of a pixel] Since the difference of brightness is large, energy $C(m, s)$ about brightness becomes large too much, right evaluation carries out, and it is for *****.

[0097] For example, the case where matching of people's face and the face of a cat is taken is considered. The face of a cat is covered with hair and the very bright pixel and the very dark pixel are mixed. In this case, in order to calculate the submap between two faces, a subimage is normalized first. That is, the brightness of the darkest pixel is set as 0, the brightest it is set as 255, and it asks for the brightness of other pixels by linear interpolation.

[0098] [1.7] Use how count advances to a linear according to the scan of an implementation starting point image inductive. The value of $f(m, s)$ is determined about pixel [of the left end on introduction and No. 1] $(i, j) = (0, 0)$. Next, $f(m, s)(i, j)$ value each is determined, increasing i every [1]. When the value of i reaches the width of face of an image, the value of j is increased one and i is returned to 0. Henceforth, $f(m, s)(i, j)$ is determined with the scan of a starting point image. If correspondence of a pixel is decided about all points, one map $f(m, s)$ will be decided.

[0099] If corresponding points $q_f(i, j)$ are decided about certain $p(i, j)$, the corresponding points $q_f(i, j+1)$ of $p(i, j+1)$ will be decided next. Under the present circumstances, the location of $q_f(i, j+1)$ is restricted by the location of $q_f(i, j)$ in order to fulfill bijection conditions. Therefore, in the point that corresponding points are decided previously, in this system, a priority becomes high. If the condition that priority is always (0 0) given most continues, an excessive deviation will join the last map called for. With this premise technique, in order to avoid this condition, $f(m, s)$ is decided by the following approaches.

[0100] When it is 0 first (second mod 4), it decides making (0, 0) into a start point and increasing i and j gradually. When (second mod 4) is 1, the right end point of the top line is made into a start point, and i is decided, making reduction and j increase. When (second mod 4) is 2, the right end point of the lowest line is made into a start point,

and it decides, decreasing i and j. When (s mod 4) is 3, the left end point of the lowest line is made into a start point, and i is decided, decreasing an increment and j. Since the concept s of a submap, i.e., a parameter, did not exist in the n-th level with the finest resolution, even if it was s= 0 and s= 2, two directions were calculated continuously.

[0101] In the actual implementation, the value of f (m, s) (i, j) (m= 0, ..., n) which fulfills bijection conditions as much as possible out of a candidate (k, l) was chosen by giving a penalty to the candidate who breaks bijection conditions. phi is hung on the energy D of the candidate who breaks the 3rd condition (k, l), and, on the other hand, psi is hung on the candidate who breaks the 1st or 2nd condition. phi= 2 and psi= 100000 were used this time.

[0102] The following tests were performed when determining =f (m, s) (i, j) as a procedure actual for the check of the above-mentioned bijection conditions (k, l). That is, it confirms whether z component of the outer product of a degree type becomes zero or more to each lattice point (k, l) contained in the inheritance quadrilateral of f (m, s) (i, j).

[0103]

[Equation 39]

$$W = \vec{A} \times \vec{B} \quad (式 3 9)$$

However, it is here and is [Equation 40].

$$\vec{A} = \overrightarrow{q_{f(m,s)(i,j-1)}^{(m,s)} q_{f(m,s)(i+1,j-1)}^{(m,s)}} \quad (式 4 0)$$

[Equation 41]

$$\vec{B} = \overrightarrow{q_{f(m,s)(i,j-1)}^{(m,s)} q_{(k,l)}^{(m,s)}} \quad (式 4 1)$$

Come out and it is (a vector is made into a three-dimensions vector here, and the z-axis is defined in a rectangular right-handed coordinate system). If W is negative, it will be made whether a penalty is given and is made and not to choose by hanging psi on D (m, s) (k, l) about the candidate.

[0104] Drawing 5 (a) and drawing 5 (b) show the reason for inspecting this condition. The candidate to whom drawing 5 (a) does not have a penalty, and drawing 5 (b) express a candidate with a penalty, respectively. In case the map f (m, s) to a

contiguity pixel (i, j+1) (i, j+1) is determined, if z component of W is negative, the pixel which satisfies bijection conditions on a starting point image plane does not exist. Because, q (m, s) (k, l) is for exceeding the boundary line of the adjoining quadrilateral.

[0105] [1.7.1] In the sequence implementation of a submap, when the number of resolution level was even, sigma(0) =3, sigma(1) =2, sigma(2) =1, sigma(3) =0, and sigma(4) =3 were used using sigma(0) =0, sigma(1) =1, sigma(2) =2, sigma(3) =3, and sigma(4) =0 at the time of odd number. By this, the submap was shuffled moderately. In addition, originally a submap is four kinds and s is either 0-3. However, processing which is equivalent to s= 4 in fact was performed. The reason is mentioned later.

[0106] [1.8] After the map between a interpolation count starting point image and a terminal point image is determined, the brightness which is the pixel which corresponds and suits is interpolated. Try linear interpolation was used in the experiment. It is assumed that square p(i, j) p(i+1, j) p(i, j+1) p (i+1, j+1) in a starting point image plane is projected on quadrilateral qf(i, j) qf(i+1, j) qf(i, j+1) qf on a terminal point image plane (i+1, j+1). Since it is easy, distance between images is set to 1. The pixel r (x y, t) of the middle image whose distance from a starting point image plane is t (0<=t<=1) (0 <=x<=N -1, 0 <=y<=M -1) is called for in the following ways. It asks for the location (however, x, y, t**R) of Pixel r (x y, t) by the degree type first.

[0107]

[Equation 42]

$$\begin{aligned} (x, y) = & (1 - dx)(1 - dy)(1 - t)(i, j) + (1 - dx)(1 - dy)tf(i, j) \\ & + dx(1 - dy)(1 - t)(i + 1, j) + dx(1 - dy)tf(i + 1, j) \\ & + (1 - dx)dy(1 - t)(i, j + 1) + (1 - dx)dytf(i, j + 1) \\ & + dxdy(1 - t)(i + 1, j + 1) + dxdytf(i + 1, j + 1) \end{aligned}$$

(式 4 2)

The brightness of the pixel in r (x y, t) is continuously determined using the following formula.

[0108]

[Equation 43]

$$\begin{aligned} V(r(x, y, t)) = & (1 - dx)(1 - dy)(1 - t)V(p_{(i,j)}) + (1 - dx)(1 - dy)tV(q_{f(i,j)}) \\ & + dx(1 - dy)(1 - t)V(p_{(i+1,j)}) + dx(1 - dy)tV(q_{f(i+1,j)}) \\ & + (1 - dx)dy(1 - t)V(p_{(i,j+1)}) + (1 - dx)dytV(q_{f(i,j+1)}) \\ & + dxdy(1 - t)V(p_{(i+1,j+1)}) + dxdytf(q_{f(i+1,j+1)}) \end{aligned}$$

(式 4 3)

dx and dy are parameters and change from 0 to 1 here.

[0109] [1.9] The decision of a map in case a constraint does not exist at all until now

[map] when imposing a constraint was described. However, when correspondence relation is beforehand specified between the pixels of the specification of a starting point image and a terminal point image, a map can be determined after making this into a constraint.

[0110] By the rough map which moves the specific pixel of a starting point image to the specific pixel of a terminal point image first, a fundamental idea transforms a starting point image roughly, and calculates Map f correctly after an appropriate time.

[0111] It opts for the rough map which projects the specific pixel of a starting point image on the specific pixel of a terminal point image, and projects other pixels of a starting point image on a suitable location first. That is, the pixel near a specific pixel is the map which is projected near the location where the specific pixel is projected. The rough map of the m-th level is described to be F (m) here.

[0112] It opts for the rough map F in the following ways. First, a map is specified about some pixels. They are the pixel of ns individual, and [Equation 44] about a starting point image.

$$P(i_0, j_0), P(i_1, j_1), \dots, P(i_{n_s-1}, j_{n_s-1}) \quad (式 4 4)$$

The following values are decided when specified.

[Equation 45]

$$\begin{aligned} F^{(n)}(i_0, j_0) &= (k_0, l_0), \\ F^{(n)}(i_1, j_0) &= (k_1, l_1), \dots, \\ F^{(n)}(i_{n_s-1}, j_{n_s-1}) &= (k_{n_s-1}, l_{n_s-1}) \end{aligned}$$

(式 4 5)

The amount of displacement of other pixels of a starting point image is an average called for by making weighting the variation rate of p (ih, jh) (h= 0, --, ns-1). That is, Pixel p (i, j) is projected on the following pixels of a terminal point image.

[0113]

[Equation 46]

$$F^{(m)}(i, j) = \frac{(i, j) + \sum_{h=0}^{n_s-1} (k_h - i_h, l_h - j_h) weight_h(i, j)}{2^{n-m}} \quad (式 4 6)$$

However, it is here and is [Equation 47].

$$weight_k(i, j) = \frac{1/||i_k - i, j_k - j||^2}{total\ weight(i, j)} \quad (式\ 4\ 7)$$

[Equation 48]

$$total\ weight(i, j) = \sum_{h=0}^{h=n_s-1} 1/||i_h - i, j_h - j||^2 \quad (式\ 4\ 8)$$

It carries out.

[0114] It continues, and the energy D of the map f (m, s) (i, j) is changed so that the candidate map f near F (m) may have fewer energy. Correctly, D (m, s) (i, j) is [Equation 49].

$$D_{(i,j)}^{(m,s)} = E_{0(i,j)}^{(m,s)} + \eta E_{1(i,j)}^{(m,s)} + \kappa E_{2(i,j)}^{(m,s)} \quad (式\ 4\ 9)$$

It comes out. However, [Equation 50]

$$E_{2(i,j)}^{(m,s)} = \begin{cases} 0, & \text{if } ||F^{(m)}(i, j) - f^{(m,s)}(i, j)||^2 \leq \lfloor \frac{\rho^2}{2^{2(n-m)}} \rfloor \\ ||F^{(m)}(i, j) - f^{(m,s)}(i, j)||^2, & \text{otherwise} \end{cases} \quad (式\ 5\ 0)$$

It comes out, and it is and is referred to as kappa and rho>=0. f is completely determined as the last according to the automatic calculation process of the above-mentioned map.

[0115] Here, f (m, s) (i, j) is at the time sufficiently near F (m), and (i, j) that is, and those distance is [Equation 51].

$$\left[\frac{\rho^2}{2^{2(n-m)}} \right] \quad (式\ 5\ 1)$$

When it is less than, it should be cautious of E2 (m, s) (i, j) being set to 0. As long as f (m, s) (i, j) is close each to F (m), and (i, j) enough, the reason defined such is because he wants to decide the value automatically, as it settles in a suitable location in a terminal point image. It is automatically mapped so that it is not necessary to specify exact correspondence relation as a detail and a starting point image may match a terminal point image for this reason.

[0116] [2] Explain the flow of processing by each component engineering of concrete procedure [1]. Drawing 6 is a flow chart which shows the whole premise technical procedure. As shown in this drawing, processing which used the multiplex resolution singular point filter first is performed (S1), and matching of a starting point image and a terminal point image is taken continuously (S2). However, S2 may process image recognition etc. based on the description of the image obtained by S1 rather than is

indispensable.

[0117] Drawing 7 is a flow chart which shows the detail of S1 of drawing 6. Here, it is premised on taking matching of a starting point image and a terminal point image by S2. Therefore, first, with a singular point filter, a starting point image is hierarchized (S10) and a series of starting point hierarchy images are obtained. A terminal point image is continuously hierarchized by the same approach (S11), and a series of terminal point hierarchy images are obtained. However, the sequence of S10 and S11 is arbitrary, and can also generate a starting point hierarchy image and a terminal point hierarchy image in parallel.

[0118] Drawing 8 is a flow chart which shows the detail of S10 of drawing 7. Size of the starting point image of a basis is set to $2n \times 2n$. Since a starting point hierarchy image is made sequentially from the one where resolution is finer, the parameter m which shows the resolution level set as the object of processing is set to n (S100). It continues. The images $p(m, 0)$, $p(m, 1)$, and $p(m, 2)$ of the m -th level, The singular point is detected using a singular point filter from $p(m, 3)$ (S101), and the images $p(1 \ m-0)$, $p(1 \ m-1)$, $p(1 \ m-2)$, and $p(1 \ m-3)$ of the m -1st level are generated, respectively (S102). Here, since it is $m=n$, it is $p(m, 0) = p(m, 1) = p(m, 2) = p(m, 3) = p(n)$, and four kinds of subimages are generated from one starting point image.

[0119] Drawing 9 shows the correspondence relation between some images of the m -th level, and some images of the m -1st level. The numeric value of this drawing shows the brightness of each pixel. In $p(m, s)$ of this drawing symbolizing four images of $p(m, 0)$ – $p(m, 3)$ and generating $p(1 \ m-0)$, $p(m, s)$ thinks that it is $p(m, 0)$. Under the regulation shown by [1.2], $p(1 \ m-0)$ about the block as which brightness was filled in for example, in this drawing “3” and $p(1 \ m-1)$ replace “8” and $p(1 \ m-2)$ among 4 pixels contained there by one pixel which acquired “10” for “6” and $p(1 \ m-3)$, respectively, and acquired this block, respectively. Therefore, the size of the subimage of the m -1st level is set to $2^{m-1} \times 2^{m-1}$.

[0120] The decrement of the m is carried out continuously (S103 of drawing 8), and it checks that m is not negative (S104), it returns to S101, and then the subimage with coarse resolution is generated. As a result of this repeat processing, when $m=0$, i.e., the subimage of the 0th level, is generated, S10 is completed. The size of the subimage of the 0th level is 1×1 .

[0121] Drawing 10 has illustrated the starting point hierarchy image generated by S10 about the case of $n=3$. Only the first starting point image is common to four sequences, and the subimage is henceforth generated independently according to the class of singular point, respectively. In addition, processing of drawing 8 is common also to S11 of drawing 7, and a terminal point hierarchy image is also generated through the same procedure. Processing by S1 of drawing 6 is completed above.

[0122] With a premise technique, in order to progress to S2 of drawing 6, matching evaluation is prepared. Drawing 11 shows the procedure. As shown in this drawing, two

or more valuation plans are set up first (S30). energy $C(m, s)$ f about the pixel introduced by [1.3.2.1], and energy $D(m, s)$ f about the smoothness of the map introduced by [1.3.2.2] -- then, it is. Next, these valuation plans are unified and a comprehensive valuation plan is stood (S31). gross energy $\lambda C(m, s) f + D(m, s) f$ introduced by [1.3.2.3] -- then -- if η which it is and was introduced by [1.3.2.2] is used $\sum (\lambda C(m, s) (i, j) + \eta E0(m, s) (i, j) + E1 (i, j) (m, s))$ (formula 52) It becomes. However, total calculates by 0, 1 --, and $2m-1$ about i and j , respectively. Preparation of the matching evaluation by the above is completed.

[0123] Drawing 12 is a flow chart which shows the detail of S2 of drawing 6 . As [1] described, matching of a starting point hierarchy image and a terminal point hierarchy image is mutually taken by the images of the same resolution level. In order to take global matching between images good, resolution calculates matching sequentially from coarse level. Since the starting point hierarchy image and the terminal point hierarchy image are generated using a singular point filter, the location and brightness of the singular point are clearly saved also on level with coarse resolution, and the result of global matching becomes what was very excellent compared with the former.

[0124] Like drawing 12 , 0 and the level parameter m are first set as 0 for the multiplier parameter η (S20). It asks for four kinds of submaps ($s=0, 1, 2, 3$) (m, s) f which continue, calculate matching between each of four subimages of the m -th level in four subimages of the m -th level in a starting point hierarchy image, and terminal point hierarchy images, and fulfill bijection conditions, respectively, and make energy min (S21). Bijection conditions are inspected using the inheritance quadrilateral stated by [1.3.3]. Under the present circumstances, as shown in formulas 17 and 18, since the submap in the m -th level is restrained by them of the $m-1$ st level, sequential use of the matching in level with more coarse resolution is carried out. This is perpendicular reference between different level. In addition, although it is $m=0$ now and there is no level coarser than it, this exceptional processing is later mentioned by drawing 13 .

[0125] On the other hand, horizontal reference in the same level is also performed. like the formula 20 of [1.3.3] -- $f(m, 3)$ decides that $f(m, 2)$ is similar to $f(m, 1)$, and $f(m, 1)$ is similar to $f(m, 0)$ at $f(m, 2)$, respectively. If the reason is included in the starting point image with the them same from the first, and the terminal point image at all even if the class of singular point is different, it is because the situation that submaps completely differ is unnatural. Energy becomes small, so that submaps are near, as shown in a formula 20, and it considers that matching is good.

[0126] In addition, since there is no submap which can be referred to on the level same about $f(m, 0)$ which should be decided first, as shown in a formula 19, level coarse [one] is referred to. However, in the experiment, after being found to $f(m, 3)$, the procedure of updating $f(m, 0)$ once by making this into a constraint was taken. It is equal to this substituting $s=4$ for a formula 20, and setting $f(m, 4)$ to new $f(m, 0)$. It is

for avoiding the inclination for the degree of association of $f(m, 0)$ and $f(m, 3)$ to become low too much, and the experimental result became fitness more by this measure. In addition to this measure, in the experiment, the shuffle of the submap shown in [1.7.1] was also performed. It is the meaning which keeps close the degree of association of the submaps it is originally decided for every class of singular point that this will also be. Moreover, in order to avoid the deviation depending on the start point of processing, the point of changing the location of a start point according to the value of s is as [1.7] having described.

[0127] Drawing 13 is drawing showing signs that a submap is determined in the 0th level. Since each ***** merely consists of one pixel, it is altogether decided on the 0th level automatically that four submaps $f(0, s)$ will be unit maps. Drawing 14 is drawing showing signs that a submap is determined in the 1st level. A subimage is constituted from 4 pixels by the 1st level, respectively. These 4 pixels are shown by the continuous line in this drawing. The following procedures are completed when looking for the corresponding points of the point x of $p(1, s)$ in $q(1, s)$ now.

[0128] 1. Search for the upper left point a of Point x , the upper right point b , the lower left point c , and the lower right point d in the resolution of the 1st level.

2. Look for the pixel to which one point $a-d$ belongs in coarse level, i.e., the 0th level. In the case of drawing 14, point $a-d$ belongs to pixel $A-D$, respectively. However, pixel $A-C$ is an imagination pixel which originally does not exist.

3. Plot corresponding-points A' of pixel $A-D$ which is the 0th level and has already been found - D' in $q(1, s)$. Pixel $A' - C'$ shall be an imagination pixel, and shall be in the respectively same location as pixel $A-C$.

4. Consider that corresponding-points a' of the point a in Pixel A is in pixel A' , and plot point a' . At this time, it is assumed that the location which point a' occupies in pixel A' is the same as the location (lower right in this case) which Point a occupies in Pixel A . Corresponding-points $b' - d'$ is plotted by the same approach as 5.4, and an inheritance quadrilateral is made from point $a' - d'$.

6. Look for corresponding-points x' of Point x so that energy may become min in an inheritance quadrilateral. As a candidate of corresponding-points x' , the core of a pixel may limit to what is contained in an inheritance quadrilateral. In the case of drawing 14, all of four pixels become a candidate.

[0129] It is the decision procedure of the corresponding points of the point x with the above. Same processing is performed about all other points, and it opts for a submap. On with a level [2nd / more than] level, since it is thought that an inheritance quadrilateral loses its shape gradually, the situation that spacing of pixel $A' - D'$ is vacant as shown in drawing 3 occurs.

[0130] In this way, if four submaps of a certain m -th level are decided, m will be incremented (S22 of drawing 12), and it will confirm that m is not over n (S23), and will return to S21. It asks for the submap of the level of resolution gradually fine whenever

it returns to S21 hereafter, and when it finally returns to S21, map [of the n -th level] $f(n)$ is decided. Since this map becomes settled about $\eta = 0$, it writes it as $f(n)$ and ($\eta = 0$).

[0131] Only $\Delta\eta$ shifts η in order to ask also for the map about next different η , and the zero clear of the m is carried out (S24). It checks that new η is not over predetermined retrieval cancel value η_{\max} (S25), and S21 is asked for map $f(n)$ and ($\eta = \Delta\eta$) about return and this η . This processing is repeated and it asks for $f(n)$ ($\eta = \Delta\eta$) ($i = 0, 1, \dots$) by S21. When η exceeds η_{\max} , it progresses to S26, and the optimal $\eta = \eta_{\text{opt}}$ is determined by the below-mentioned approach, and, finally $f(n)$ and ($\eta = \eta_{\text{opt}}$) are set to map $f(n)$.

[0132] Drawing 15 is a flow chart which shows the detail of S21 of drawing 12. With this flow chart, the submap in the m -th level is decided about certain fixed η . In case it opts for a submap, with a premise technique, optimal λ is decided independently for every submap.

[0133] As shown in this drawing, the zero clear of s and λ is carried out first (S210). Next it asks for the submap $f(m, s)$ which makes energy min about λ at that time (it reaches and is about η to dark) (S211), and this is written to be $f(m, s)$ ($\lambda = 0$). Only $\Delta\lambda$ shifts λ in order to ask also for the map about different λ , and it checks that new λ is not over predetermined retrieval cancel value λ_{\max} (S213), and S211 is asked for $f(m, s)$ ($\lambda = \Delta\lambda$) ($i = 0, 1, \dots$) by return and subsequent repeat processings. When λ exceeds λ_{\max} , it progresses to S214, and the optimal $\lambda = \lambda_{\text{opt}}$ is determined, and, finally $f(m, s)$ ($\lambda = \lambda_{\text{opt}}$) is considered as Map $f(m, s)$ (S214).

[0134] Next, the zero clear of λ is carried out in order to ask for other submaps in the same level, and s is incremented (S215). It checks that s is not over 4 (S216), and returns to S211. If set to $s = 4$, $f(m, 0)$ will be updated like **** using $f(m, 3)$, and the decision of the submap in the level is ended.

[0135] Drawing 16 is drawing showing the behavior of energy $C(m, s)$ f corresponding to $f(m, s)$ ($\lambda = \Delta\lambda$) ($i = 0, 1, \dots$) called for while changing λ about certain m and s . If λ increases, $C(m, s)$ f will usually decrease, as [1.4] described. However, if λ exceeds an optimum value, $C(m, s)$ f will start to increase. So, with this premise technique, λ in case $C(m, s)$ f takes the minimal value is decided to be λ_{opt} . What is necessary is just to note the first minimum point, in order for the map to already have collapsed at the time and not to make semantics, even if $C(m, s)$ f becomes small again in the range of $\lambda > \lambda_{\text{opt}}$, as shown in this drawing. λ_{opt} is decided independently for every submap, and, finally becomes settled one also about $f(n)$.

[0136] On the other hand, drawing 17 is drawing showing the behavior of Energy $C(n)$ f corresponding to $f(n)$ ($\eta = \Delta\eta$) ($i = 0, 1, \dots$) called for while changing η . If η increases also here, $C < \text{SUP} > (n)$ f will usually decrease, but $C(n)$ f will start to increase

if η exceeds an optimum value. Then, η in case $C(n)$ takes the minimal value is decided to be η_{opt} . Drawing 17 may consider drawing which expanded near the zero of the axis of abscissa of drawing 4. If η_{opt} is decided, terminal decision of the $f(n)$ can be carried out.

[0137] As mentioned above, according to this premise technique, various merits are obtained. Since it is not necessary to detect an edge probably, the technical problem of the edge detection type conventional technique is cancelable. Moreover, the transcendental knowledge over the object contained in an image is also unnecessary, and automatic detection of corresponding points realizes it. According to the singular point filter, the brightness and location of the singular point can be maintained also on level with coarse resolution, and it is very advantageous to object recognition, a feature extraction, and image matching. Consequently, construction of the image processing system which mitigates a help activity sharply is attained.

[0138] In addition, the following deformation techniques are also considered about this premise technique.

(1) With the premise technique, when taking matching between a starting point hierarchy image and a terminal point hierarchy image, the automatic decision of a parameter was made, but this approach can be used for the whole when taking matching between [instead of / between hierarchy images] the usual images of two sheets.

[0139] For example, between the images of two sheets, make two, the energy E_0 about the difference of the brightness of a pixel, and the energy E_1 about a location-gap of a pixel, into a valuation plan, and let these linear combination $E_{tot} = \alpha E_0 + E_1$ be a comprehensive valuation plan. An automatic decision of the α is made paying attention to near the extremal value of this comprehensive valuation plan. That is, it asks for a map to which E_{tot} becomes min about various α . α in case E_1 takes the minimal value about α among those maps is decided to be an optimal parameter. Finally it is considered that the map corresponding to the parameter is the optimal matching between both images.

[0140] What there are various approaches in a setup of a valuation plan besides this, for example, takes such a big value like $1/E_1$, and $1/E_2$ that an evaluation result is good may be adopted. A comprehensive valuation plan does not necessarily need to be linear combination, either, and should just choose suitably the function of the n -th power sums ($n = 2, 1/2, -1, -2$, etc.), a polynomial, and arbitration etc.

[0141] In the case of two (η and λ), any, such as a case beyond it, are sufficient also as a parameter only for α like a premise technique. When [one / every] a parameter is three or more, it is made to change and decides.

[0142] (2) With this premise technique, after deciding a map that the value of a comprehensive valuation plan becomes min, the point that $C(m, s)$ f which is one valuation plan which constitutes a comprehensive valuation plan became the minimum

was detected, and the parameter was determined. However, it is effective even if it decides a parameter that the minimum value of a comprehensive valuation plan only becomes min depending on a situation instead of such two-step time processing. $\alpha E_0 + \beta E_1$ is made into a comprehensive valuation plan in that case, the constraint which becomes $\alpha + \beta = 1$ may be established and measures, such as treating each valuation plan equally, may be taken. The energy of the essence of the automatic decision of a parameter is because it is in the point of deciding the parameter to become min.

[0143] (3) The premise technique generated four kinds of subimages about four kinds of singular points on each resolution level. however -- naturally -- 4 kinds of inside 1 -- 2 or 3 kinds may be used alternatively. For example, if it is in the condition that only one bright point exists in an image, suitable effectiveness should be acquired even if it generates a hierarchy image only by $f(m, 3)$ about the maximum point. In that case, since a submap which is different on the same level becomes unnecessary, it is effective in the computational complexity about s becoming less.

[0144] (4) With this premise technique, when one level progressed with the singular point filter, the pixel was set to one fourth. For example, a pixel will be set to one ninth, if the configuration which considers as 1 block by 3×3 , and looks for the singular point in it is also possible and one level progresses in that case.

[0145] (5) When a starting point image and a terminal point image are colors, change them into monochrome image first and calculate a map. The color picture of the starting point is changed using the map called for as a result. As the other approach, a submap may be calculated about each component of RGB.

[0146] The gestalt of operation using the premise technique beyond [the image effectiveness or the gestalt of the operation about image interpolation] is explained. Although the gestalt 1 of operation mentions the thing more near application after the thing more near a principle, and the gestalt 2 of operation, it is not necessary to draw a clear boundary line to them, and one of the functions of the arbitration of the gestalt of operation and the function of the arbitration of the gestalt of other operations may be combined with them.

[0147] Gestalt 1 drawing 18 of operation shows the 1st image I1 and the 2nd image I2 which are a key frame, and some pixels $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$ correspond among them. These correspondences are called for with the premise technique.

[0148] Drawing 19 cuts a mesh on the 1st image I1, and shows the correspondence location on the 2nd image I2 of the polygon which constitutes the mesh. The polygon R1 now observed on the 1st image I1 was decided by the four lattice points A, B, C, and D. This polygon R1 is named an "origin polygon." these -- four -- a piece -- the lattice point -- A -- B -- C -- D -- drawing 18 -- having been shown -- as -- respectively -- the -- two -- an image -- I -- two -- a top -- corresponding -- a point -- A -- ' -- B -- ' -- C -- ' -- D -- ' -- having -- these -- a point -- forming

-- having -- a polygon -- R -- two -- "a terminal point polygon" -- naming . Although an origin polygon is generally a rectangle, generally a terminal point polygon is a quadrilateral. Anyway, the gestalt of this operation does not describe the correspondence relation between the 1st image I1 and the 2nd image I2 per pixel, but describes that correspondence pixel about the lattice point of an origin polygon. The description is written in a corresponding-points file. By observing the lattice point, capacity of a corresponding-points file can be made very small.

[0149] A corresponding-points file is used in order to generate the middle image of the 1st image I1 and the 2nd image I2. This point can generate the middle image of the time location of arbitration by interpolating the location of corresponding points, as the premise technique described. Therefore, morphing and the smooth animation between two images are generable to it by saving the 1st image I1, the 2nd image I2, and a corresponding-points file. For this reason, the compression effectiveness of an animation is acquired.

[0150] Drawing 20 shows how to calculate the correspondence relation about points other than the lattice point from a corresponding-points file. Since there is only information about the lattice point in a corresponding-points file, it is necessary to calculate correspondence information separately about the point inside a polygon. Drawing 20 shows correspondence with the triangle ABC which hits the lower half of the origin polygon R1 of drawing 19 , and triangle A'B'C' which hits the lower half of the terminal point polygon R2 similarly. Suppose that it was in the location where the point Q inside triangle ABC by the side of an origin polygon divides Segment AC interiorly to t: (1-t), and divides interiorly between the internally dividing point and Point B to s: (1-s) now. What is necessary is just to think that corresponding-points Q' in triangle A'B'C' by the side of the terminal point polygon of this point Q is in the location which divides segment A'C' interiorly to t: (1-t), and divides interiorly between that internally dividing point and point B' to s: (1-s). Namely, what is necessary is to divide an origin polygon into a triangle and just to search for the point inside a terminal point polygon in the form of the internal division of the vector about the triangle. If vector italic shows, it will be $BQ = (1-s) \{ (1-t) BA + tBC \}$.

coming out -- it is -- therefore, $B'Q' = (1-s) \{ (1-t) B'A' + tB'C' \}$

Though natural, suppose that the same processing is made also about between the triangle ACD which hits the upper half of the origin polygon R1, and triangle A'C'D' which hits the upper half of the terminal point polygon R2 similarly.

[0151] Drawing 21 shows the above procedure. The matching result about the lattice point taken on the 1st image I1 is first acquired like drawing 19 (S10). What is necessary is to perform matching of the pixel unit of a premise technique and just to extract the part which hits at the lattice point out of that result at this time. However, the matching result about the lattice point may be specified others, a matching technique, for example, an optical flow, and based on block matching, without using a

premise technique.

[0152] It continues and a terminal point polygon is defined on the 2nd image I2 like the right of drawing 19 (S12). Since a corresponding-points file is above generable, the information for identifying the 1st image I1 and the 2nd image I2 is incorporated and outputted to this file (S14). Two images and a corresponding-points file are held at the recording device or medium of arbitration, or are put on a direct network or a broadcast wave, and should just be transmitted.

[0153] Drawing 22 shows the procedure which generates a middle image using a corresponding-points file. First, the 1st image I1 and the 2nd image I2 are read (S20), and a corresponding-points file is read (S22). It continues and the correspondence relation between the point in an origin polygon and the point in a terminal point polygon is computed by the approach of drawing 20 (S24). At this time, the correspondence relation about all the pixels in an image is acquirable. Therefore, if the coordinate and color of a point which correspond each other are divided interiorly by u : $(1-u)$ as the premise technique described, the middle image in the location which divides interiorly between the 1st image I1 and the 2nd image I2 by u in time $(1-u)$ is generable (S26). In addition, not only interpolation interpolation but extrapolation interpolation may be carried out.

[0154] Drawing 23 shows the configuration of the image effectiveness equipment 10 which performs the above processing. The image input section 12 in which this equipment 10 acquires the 1st image I1 and the 2nd image I2 from external storage, an external photography camera, etc., The matching processor 14 which performs matching count to those images with the technique of a premise technique and others, The corresponding-points file attaching part 16 which stores the corresponding-points file F generated by it, The middle image generation section 18 which generates a middle image from the 1st image I1 and the 2nd image I2, and the display 20 displayed as an image near the animation of while carrying out timing adjustment of the 1st image I1, the 2nd image I2, and the middle image are included. Moreover, the communications department 22 sends out the 1st image I1, the 2nd image I2, and the corresponding-points file F to the transmission infrastructure of a network and others according to the request from the outside. In addition, the mesh information which shows the magnitude of a mesh, the location of the lattice point, etc. is inputted into the matching processor 14.

[0155] By the above configuration, the 1st image I1 and the 2nd image I2 which were inputted into the image input section 12 are sent to the matching processor 14. The matching processor 14 performs matching count of a pixel unit among those images. The matching processor 14 generates the corresponding-points file F based on mesh information, and outputs it to the corresponding-points file attaching part 16.

[0156] the middle image generation section 18 -- a user -- it responds for asking, or the corresponding-points file F is read by other factors, and a middle image is

generated. This middle image is sent to a display 20, timing of an image output is performed there, and an animation or a morphing image is displayed. This equipment 10 may be in the remote terminal side of another configuration, that terminal receives the 1st image I1, the 2nd image I2, and comparatively light data called the corresponding-points file F in that case, and the middle image generation section 18 and a display 20 can reproduce an animation independently as this actuation also shows.

[0157] The communications department 22 is a configuration for it, and it assumes that the remote terminal has already waited at the network point here. That is, the communications department 22 puts and sends out the 1st image I1, the 2nd image I2, and the corresponding-points file F to a network or a broadcast wave, and the display of an animation is performed by the remote terminal side. Of course, a remote terminal is good for the purpose of not a display but storage. In addition, this equipment inputs the 1st image I1, the 2nd image I2, and those corresponding-points files through the communications department 22 from the exteriors, such as a network, passes these to the middle image generation section 18, carries out interpolation count, and also has the application which generates a middle image. The path P of the data for it is described by drawing 23 .

[0158] When experimented about the contents of processing of the gestalt 1 of operation, very good morphing or the animation compression effectiveness was acquired by adopting about 256x256 magnitude and setting the lattice point as every direction 10 – every dozens of pixel extent as the 1st and 2nd image. The magnitude of a corresponding-points file is about 10 K bytes from several K, and it has checked that the height of image quality and the smallness of the amount of data could be realized.

[0159] The image effectiveness equipment realized with the gestalt 1 of gestalt 2 implementation of operation can be used for the goods presentation in electronic commerce. This equipment (henceforth “goods presentation equipment”) of the gestalt 2 of operation uses as the 1st image I1 and the 2nd image I2 the image of two sheets which approaches among the photographs of the goods photoed from some include angles, and suits, and goods are shown to a user by interpolating between them smoothly. For example, a rotation image smooth 360 degrees is obtained by carrying out sequential interpolation of the photograph taken from the surroundings of goods at intervals of 10 degrees. Conventionally, the smooth rotation image of goods is using the technique of the gestalt 1 of operation, although photography was needed every 1 or 2 degrees at most, and it has become clear that this include angle is made to 10–20 degrees.

[0160] Unlike the gestalt 1 of operation, goods presentation equipment uses the corresponding-points file F generated beforehand. The photograph of goods is because it can prepare before presentation. If only difference with the configuration of

drawing 23 is shown, the image input section 12 will acquire the corresponding-points file F other than the 1st image I1 and the 2nd image I2. These are inputted into the middle image generation section 18, and a middle image is generated. A middle image may be displayed by the display 20, or may be sent out through the communications department 22 to the transmission infrastructure of a network and others. With the gestalt of this operation, the matching processor 14 and the corresponding-points file attaching part 16 are unnecessary. In addition, such a configuration is effective also in the gestalt of other operations.

[0161] In addition to the gestalt 1 of operation, goods presentation equipment offers the goods presentation function in electronic commerce. When this function is the above-mentioned goods rotation function, goods may be rotated according to the directions inputted through GUI from the user. The approach of such directions is learned [that it is various and] for the field of CG, and uses such a technique here.

[0162] The middle image which interpolated the goods image in the two directions and was obtained as another example of a function may be displayed. For example, it can interpolate not only to a longitudinal direction but to a lengthwise direction by photoing goods at intervals in addition to taking a photograph at intervals also in a lengthwise direction in a longitudinal direction. This function includes a lower process. (1) The image pair B which consists of the 3rd image of a process (2) which acquires the corresponding-points information A between the images of these [the image pair A which consists of the 1st image and the 2nd image, and] two sheets, and the 4th image The physical relationship of the shaft A which becomes settled in time or spatially between the images of two sheets of the process (3) image pair A which acquire the corresponding-points information B between the images of these two sheets, and the shaft B which becomes settled in time or spatially between the images of two sheets of the image pair B, The process which generates a middle image with interpolation using the corresponding-points information A and the corresponding-points information B.

[0163] In (3), biaxial interpolation (bilinear interpolation) may be performed using Shaft A and Shaft B. As an example, it considers as the image pair A obtained from two views p1 (0 0) and p2 (0,100), and considers as the image pair B obtained from two another views p3 (100 0) and p4 (100,100). The straight line to which the straight line which connects points p1 and p2 connects Shaft A and points p3 and p4 hits Shaft B.

[0164] When obtaining the image seen from view p' = (50 50) as a middle image, the frame from a view (0 50) is first generated based on the corresponding-points information A between the image pairs A. Next, another frame from a view (100 50) is generated based on the corresponding-points information B between the image pairs B. It continues, and interpolates on these two frames, namely, divides interiorly by 1:1 in this case, and a desired intermediate frame is generated. What is necessary is here, just to set the image pair A and the image pair B that Shaft A and Shaft B do not

come on the same straight line generally, in order to interpolate both directions in every direction.

[0165] In addition, although Shaft A and Shaft B become settled spatially between the images of two sheets in this example, respectively, there is also another example which becomes settled in time. For example, two images obtained from one view P to time-of-day $t=t_0$ and t_1 are made into the image pair A, and two images obtained from another view Q to time-of-day $t=t_0$ and t_1 are made into the image pair B. In this case, the image pair A -- setting (P, t_0) -- the straight line which connects the point defined and the point defined by (P, t_1) -- Shaft A -- becoming -- the same -- the image pair B -- setting (Q, t_0) -- the straight line which connects the point defined and the point defined by (Q, t_1) serves as Shaft B. Therefore, what is necessary is just to interpolate these, respectively, after generating the image from a point $(P+Q)/2$, $(t_0+t_1)/2$, then the in-between image of two shafts for a middle image.

[0166] It is in charge of generation of the gestalt 3 middle image of operation, and the function about the image effectiveness may be offered. Although the gestalt 1 of operation described the image effectiveness as generalities, the image effectiveness here may be considered to be a near application more concretely or professionally. It is [following] the example of a function.

[0167] Out of three or more images, two adjoining images are selected as the 1st image and the 2nd image, and those middle images are generated and displayed. It continues, another adjoining image is used as the 1st image and the 2nd image, and a middle image is generated in order henceforth. For example, when there are images S, T, and U with which views differ little by little, a series of images can be continuously displayed by making the middle image of Images S and T first, and then making the middle image of Images T and U. That is, a panoptic image will be obtained, if between many discrete view images is interpolated one after another and a middle image is generated.

[0168] Although it is the same approach as this, the element of time amount can also be taken in. For example, if two or more discrete images are caught with the image to which the view was moved with time amount and order is displayed for those middle images later on, the so-called walk-through image will be obtained.

[0169] Another function displays the middle image which interpolated the image set as the object of processing in the two directions, and was obtained. 2 shaft interpolation realizes this approach as the gestalt 2 of operation also described it. Thereby, the panorama image which carries out continuation change is also obtained for example, in the two directions.

[0170] Still more nearly another function is a function which photos a certain object with the surroundings little by little, and generates those sequential images. Although the goods itself were rotated with the gestalt 2 of operation, fun is in the point which displays the image which photoed the object round rather as image effectiveness.

Although the image which goes around the person who remained in the air is generated by the SFX technique on some movies, according to the gestalt 1 of operation, such an image is easily generable with fewer photographs.

[0171] The existing image effectiveness can be increased the efficiency of or high-definition-ized with the technique concerning the gestalt of operation besides these. The equipment concerning the gestalt of this operation may carry GUI as which a user specifies whether which image effectiveness is performed, and may mount the function of suitable pretreatment or after treatment in the contents of assignment.

[0172] For example, when generating a walk-through image, the moving trucking of a view may be specified beforehand and a middle image may be displayed in order according to it. When interpolating in the two directions, the view moved according to GUI, for example, the migration direction of a mouse, is assumed like the gestalt 2 of operation, a interpolation image is generated, and a potato is good. If a mouse moves to the upper left from the lower right, it assumes that the view also moved to the upper left from the lower right, and it should just carry out 2 shaft interpolation.

[0173] As the gestalt 1 of gestalt 4 implementation of operation also described, this invention is fit also for compression of an animation. Since an animation is reproducible based on few static images, it is advantageous in case an animation is distributed by the Internet etc. For this reason, it is in charge of generation of a middle image, and the utility feature about generation of a dynamic image may be offered.

[0174] As an example of the function, the number of sheets of the middle image which should be generated is specified. For example, if the 1st image and the 2nd image are photoed with spacing of 1 second from the first, there are usually about 30 frames among them. If a user specifies it as "30 middle images" through GUI etc., the same thing as the animation of a basis will be obtained as a time scale. If "3000 sheets" is specified here, the slow motion image which extended time amount 100 times will be obtained. Since it will end with this invention if the step size of interpolation count is only changed however large this number of sheets may become, there is no part completely difficult for processing. A camera to the extent that 3000 frames can be photoed in current and 1 second does not exist, but according to the gestalt of this operation, the super slow motion image of the time amount rate of arbitration is obtained.

[0175] Since gestalt 5 this invention of operation reduces the amount of data and can generate an animation, generally its constraint is advantageous also to the application to common mobile computing devices to a CPU power, a battery life, and memory space. Therefore, generation of a middle image may be performed under the function on condition of application to mobile computing devices.

[0176] As an example of the function, the middle image generated based on one side and said correspondence information on the 1st image or the 2nd image may be displayed. In that case, there is little capacity of the memory for setting an image

temporary, and it ends. What is necessary is just to move the image element (for it to be made to only represent with a pixel below) of the polygon and others on each pixel and the mesh of the 1st image according to correspondence information, when only correspondence information and the 1st image generate a middle image. That is, although both the location of a pixel and a pixel value, i.e., a color, can usually be interpolated for correspondence information, this is simplified to processing of only migration of the location of a pixel.

[0177] As another function, streaming processing of the 1st image and the correspondence information may be carried out, and a middle image may be generated. As an example, the middle image of $t=0-1$ is generated and displayed from image [of $1.t=0$] $I(0)$, and correspondence information. The streaming display of an animation is realized by comparatively small memory space by generating and displaying the middle image of $t=1-2$ from image [of $2.t=1$ which throws away image $I(0)$] $I(1)$, and correspondence information, and repeating processing in which image $I(1)$ is thrown away. As for a middle image, in any case, it should exist in a frame buffer only at the moment of a display.

[0178] In the above, this invention was explained based on the gestalt of some operations. Each of these functions or processings can be offered in the form of a computer program. Moreover, the combination of the arbitration of the above processing and recombination are also effective as a mode of this invention.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The image and drawing 1 (c) which are obtained with drawing 1 (a) by drawing 1 (b) giving an equalization filter to two persons' person's face, and drawing 1 (d) The image, drawing 1 (e), and drawing 1 (f) of p (5 0) which are called for with a premise technique about two persons' person's face The image, drawing 1 (g), and drawing 1 (h) of p (5 1) which are called for with a premise technique about two persons' person's face The image, drawing 1 (i), and drawing 1 (j) of p (5 2) called for with a premise technique about two persons' person's face are the photograph of the halftone image which displayed on the display the image of p (5 3) called for with a premise technique about two persons' person's face, respectively.

[Drawing 2] Drawing in which drawing 2 (R) shows the quadrilateral of a basis, drawing 2 (A), drawing 2 (B), drawing 2 (C), drawing 2 (D), and drawing 2 (E) are drawings showing an inheritance quadrilateral, respectively.

[Drawing 3] It is drawing showing the relation between a starting point image and a terminal point image, and the relation between the m -th level and the $m-1$ st level

using an inheritance quadrilateral.

[Drawing 4] It is drawing showing the relation between Parameter η and Energy Cf.

[Drawing 5] a ***** [that, as for drawing 5 (a) and drawing 5 (b), the map about a certain point fulfills bijection conditions] -- an outer product -- it is drawing showing signs that it asks from count.

[Drawing 6] It is the flow chart which shows the whole premise technical procedure.

[Drawing 7] It is the flow chart which shows the detail of S1 of drawing 6 .

[Drawing 8] It is the flow chart which shows the detail of S10 of drawing 7 .

[Drawing 9] It is drawing showing the correspondence relation between some images of the m -th level, and some images of the $m-1$ st level.

[Drawing 10] It is drawing showing the starting point hierarchy image generated with the premise technique.

[Drawing 11] Before progressing to S2 of drawing 6 , it is drawing showing the procedure of preparation of matching evaluation.

[Drawing 12] It is the flow chart which shows the detail of S2 of drawing 6 .

[Drawing 13] It is drawing showing signs that a submap is determined in the 0th level.

[Drawing 14] It is drawing showing signs that a submap is determined in the 1st level.

[Drawing 15] It is the flow chart which shows the detail of S21 of drawing 12 .

[Drawing 16] It is drawing showing the behavior of energy $C(m, s)$ f corresponding to $f(m, s)$ ($\lambda = \delta\lambda$) called for while changing λ about certain $f(m, s)$.

[Drawing 17] It is drawing showing the behavior of Energy $C(n)$ f corresponding to $f(n)$ ($\eta = \delta\eta$) ($i = 0, 1, \dots$) called for while changing η .

[Drawing 18] It is drawing showing signs that the pixel which it is between the 1st image and the 2nd image corresponds.

[Drawing 19] It is drawing showing the correspondence relation between the origin polygon taken on the 1st image, and the terminal point polygon taken on the 2nd image.

[Drawing 20] It is drawing showing the procedure of searching for the point in the terminal point polygon corresponding to the point in an origin polygon.

[Drawing 21] It is the flow chart which shows the procedure which generates a corresponding-points file with the gestalt 1 of operation.

[Drawing 22] It is the flow chart which shows the procedure which generates a middle image based on a corresponding-points file with the gestalt 1 of operation.

[Drawing 23] It is the block diagram of the image effectiveness equipment concerning the gestalt 1 of operation.

[Description of Notations]

10 The Image Effectiveness Equipment

12 Image Data Input Section

14 Matching Processor

16 Corresponding-Points File Attaching Part

18 Middle Image Generation Section

20 Display

22 Communications Department

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2002-230572
(P2002-230572A)

(43)公開日 平成14年8月16日(2002.8.16)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード*(参考)
G 0 6 T 13/00		G 0 6 T 13/00	A 5 B 0 5 0
5/20		5/20	A 5 B 0 5 7
7/00	3 0 0	7/00	3 0 0 B 5 B 0 8 0
7/20		7/20	C 5 C 0 2 3
15/00	3 0 0	15/00	3 0 0 5 L 0 9 6

審査請求 有 請求項の数44 O L (全 30 頁) 最終頁に続く

(21)出願番号 特願2001-2070(P2001-2070)
(22)出願日 平成13年1月10日(2001.1.10)
(31)優先権主張番号 特願2000-360640(P2000-360640)
(32)優先日 平成12年11月28日(2000.11.28)
(33)優先権主張国 日本(J P)

(71)出願人 396001980
株式会社モノリス
東京都港区麻布十番1丁目7番3号
(72)発明者 品川 嘉久
東京都江戸川区西葛西5-10-26
(72)発明者 永嶋 浩樹
千葉県千葉市稲毛区天台4-2-12
(72)発明者 秋吉 仰三
東京都港区麻布十番1-7-3 株式会社
モノリス内
(74)代理人 100105924
弁理士 森下 賢樹

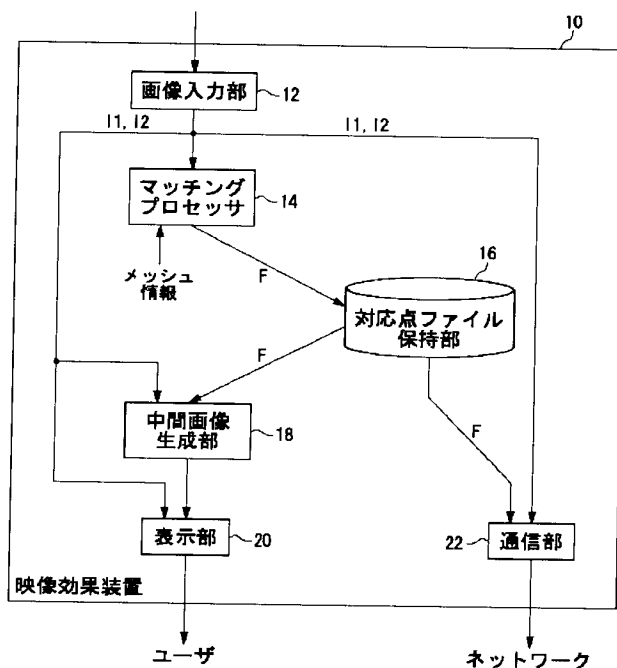
最終頁に続く

(54)【発明の名称】 映像効果方法と装置および画像補間方法と装置

(57)【要約】

【課題】 映像効果を簡便に、高い品質で、しかも軽いデータで実現することが望まれつつ、困難な要請だった。

【解決手段】 画像入力部12は第1画像と第2画像を入力する。マッチングプロセッサ14はそれらの画像間で画素マッチングを行い、第1画像上に取られたメッシュの格子点について、その対応点を第2画像上に求める。その結果が対応点ファイルFとして記録される。中間画像生成部18は、対応点ファイルFをもとに第1画像と第2画像間の中間画像を生成する。格子点だけで対応点ファイルを作ったので、そのデータ容量は非常に小さくなり、簡便に各種映像効果、動画圧縮、オンライン商品紹介、携帯機器への実装が実現する。



【特許請求の範囲】

【請求項1】 第1画像上に設けられたメッシュの格子点に対応する第2画像上の点を画像マッチングによって検出する工程と、

その結果をもとに、第1画像上のメッシュを構成する起点ポリゴンに、第2画像上にて対応する終点ポリゴンを定義する工程と、
を含むことを特徴とする映像効果方法。

【請求項2】 前記検出する工程は、第1画像と第2画像の間で画素単位のマッチング計算を行う請求項1に記載の方法。

【請求項3】 前記検出する工程は、第1画像について二次元的な探索を行って検出した特異点と、第2画像について二次元的な探索を行って検出した特異点との対応をもとに画素単位のマッチング計算を行う請求項2に記載の方法。

【請求項4】 前記検出する工程は、第1画像と第2画像を、それぞれ前記特異点を抽出することによって多重解像度化したうえで、同一解像度レベル間で画像単位のマッチング計算を行い、その結果を異なる解像度レベルにおけるマッチング計算に継承しながら、最終的に最も解像度の細かいレベルにおける画素単位の対応関係を取得する請求項3に記載の方法。

【請求項5】 前記定義する工程は、起点ポリゴンと終点ポリゴンの対応関係を対応点ファイルに出力し、このファイルを第1画像および第2画像と関連づけて保持する請求項1から4のいずれかに記載の方法。

【請求項6】 前記対応点ファイルを利用し、第1画像と第2画像の中間画像をポリゴン単位の補間計算にて生成する工程をさらに含む請求項5に記載の方法。

【請求項7】 前記生成する工程は、起点ポリゴン内部の点に対応する終点ポリゴン内部の点を導出する請求項6に記載の方法。

【請求項8】 起点ポリゴンを三角形に分割し、各三角形に関するベクトルの内分にて終点ポリゴン内部の点を求める請求項7に記載の方法。

【請求項9】 第1画像および第2画像を取得する画像入力部と、
第1画像と第2画像のマッチングを計算するマッチングプロセッサとを含み、
このマッチングプロセッサは、第1画像上のメッシュを構成する起点ポリゴンに、第2画像上にて対応する終点ポリゴンを定義することにより、対応点ファイルを生成することを特徴とする映像効果装置。

【請求項10】 前記マッチングプロセッサは、第1画像について二次元的な探索を行って検出した特異点と、第2画像について二次元的な探索を行って検出した特異点との対応をもとに画素単位のマッチング計算を行う請求項9に記載の装置。

【請求項11】 前記マッチングプロセッサは、第1画

像と第2画像を、それぞれ前記特異点を抽出することによって多重解像度化したうえで、同一解像度レベル間で画像単位のマッチング計算を行い、その結果を異なる解像度レベルにおけるマッチング計算に継承しながら、最終的に最も解像度の細かいレベルにおける画素単位の対応関係を取得する請求項10に記載の装置。

【請求項12】 前記対応点ファイルを外部へ送出するための通信部をさらに含む請求項9から11のいずれかに記載の装置。

【請求項13】 第1画像、第2画像およびそれらの間の対応点ファイルを取得し、それらふたつの画像の中間画像を生成する中間画像生成部を含み、前記対応点ファイルが第1画像上にとられたメッシュの格子点と、それに対応する第2画像上の点の位置関係を記述していることを特徴とする映像効果装置。

【請求項14】 前記中間画像、前記第1画像、前記第2画像のタイミングを調整してそれらを表示する表示部をさらに含む請求項13に記載の装置。

【請求項15】 第1画像上に設けられたメッシュの格子点に対応する第2画像上の点を画像マッチングによって検出する工程と、

その結果をもとに、第1画像上のメッシュを構成する起点ポリゴンに、第2画像上にて対応する終点ポリゴンを定義する工程と、
をコンピュータに実行せしめることを特徴とするコンピュータプログラム。

【請求項16】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、

その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成に当たり、電子商取引における商品プレゼンテーション機能を提供する工程と、
を含むことを特徴とする画像補間方法。

【請求項17】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、

その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成に当たり、映像効果に関する機能を提供する工程と、
を含むことを特徴とする画像補間方法。

【請求項18】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、
その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、

その中間画像の生成に当たり、動画画像の生成に関するユーティリティ機能を提供する工程と、を含むことを特徴とする画像補間方法。

【請求項19】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、その中間画像の生成を、モバイル機器への適用を前提とした機能のもとで行う工程と、を含むことを特徴とする画像補間方法。

【請求項20】 前記商品プレゼンテーション機能は、ユーザからの指示にしたがって商品を回転して表示する機能を含む請求項16に記載の方法。

【請求項21】 前記商品プレゼンテーション機能は、商品画像を二方向に補間して得られた中間画像を表示する機能を含む請求項16に記載の方法。

【請求項22】 前記映像効果に関する機能は、3以上の画像の中から順次ふたつの画像を第1画像および第2画像として選定し中間画像を連続的に表示する機能を含む請求項17に記載の方法。

【請求項23】 前記映像効果に関する機能は、処理の対象となる画像を二方向に補間して得られた中間画像を表示する機能を含む請求項17に記載の方法。

【請求項24】 前記ユーティリティ機能は、生成すべき中間画像の枚数を指定する機能を含む請求項18に記載の方法。

【請求項25】 前記生成すべき中間画像の枚数に関する指定はGUIを介してユーザから取得される請求項24に記載の方法。

【請求項26】 前記モバイル機器への適用を前提とした機能は、第1画像または第2画像の一方のみと前記対応情報をもとに生成された中間画像を表示する機能である請求項19に記載の方法。

【請求項27】 第1画像または第2画像の一方のみと前記対応情報をストリーミング処理して前記中間画像が生成される請求項26に記載の方法。

【請求項28】 画像を補間する装置であって、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得し、その対応情報もとに第1画像と第2画像の中間画像を生成する中間画像生成部を含み、当該中間画像生成部は、中間画像の生成に当たり、電子商取引における商品プレゼンテーション機能を提供することを特徴とする画像補間装置。

【請求項29】 画像を補間する装置であって、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポ

リゴンとの対応情報を取得し、その対応情報もとに第1画像と第2画像の中間画像を生成する中間画像生成部を含み、

当該中間画像生成部は、中間画像の生成に当たり、映像効果に関する機能を提供することを特徴とする画像補間装置。

【請求項30】 画像を補間する装置であって、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得し、その対応情報もとに第1画像と第2画像の中間画像を生成する中間画像生成部を含み、当該中間画像生成部は、中間画像の生成に当たり、動画画像の生成に関するユーティリティ機能を提供することを特徴とする画像補間装置。

【請求項31】 画像を補間する装置であって、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得し、その対応情報もとに第1画像と第2画像の中間画像を生成する中間画像生成部を含み、当該中間画像生成部は、中間画像の表示を、モバイル機器への適用を前提とした機能のもとで行うことを特徴とする画像補間装置。

【請求項32】 前記商品プレゼンテーション機能は、ユーザからの指示にしたがって商品を回転して表示する機能を含む請求項28に記載の装置。

【請求項33】 前記商品プレゼンテーション機能は、商品画像を二方向に補間して得られた中間画像を表示する機能を含む請求項28に記載の装置。

【請求項34】 前記映像効果に関する機能は、3以上の画像の中から順次ふたつの画像を第1画像および第2画像として選定し中間画像を連続的に表示する機能を含む請求項29に記載の装置。

【請求項35】 前記映像効果に関する機能は、処理の対象となる画像を二方向に補間して得られた中間画像を表示する機能を含む請求項29に記載の装置。

【請求項36】 前記ユーティリティ機能は、生成すべき中間画像の枚数を指定する機能を含む請求項30に記載の装置。

【請求項37】 前記生成すべき中間画像の枚数に関する指定はGUIを介してユーザから取得される請求項36に記載の装置。

【請求項38】 前記モバイル機器への適用を前提とした機能は、第1画像または第2画像の一方のみと前記対応情報をもとに生成された中間画像を表示する機能である請求項31に記載の装置。

【請求項39】 第1画像または第2画像の一方のみと前記対応情報をストリーミング処理して前記中間画像が生成される請求項38に記載の装置。

【請求項40】 前記中間画像、前記第1画像、前記第2画像のタイミングを調整してそれらを表示する表示部をさらに含む請求項28から39のいずれかに記載の装置。

【請求項41】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、
その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成に当たり、電子商取引における商品プレゼンテーション機能を提供する工程と、
をコンピュータに実行せしめることを特徴とするコンピュータプログラム。

【請求項42】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、
その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成に当たり、映像効果に関する機能を提供する工程と、
をコンピュータに実行せしめることを特徴とするコンピュータプログラム。

【請求項43】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、
その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成に当たり、動画の生成に関するユーティリティ機能を提供する工程と、
をコンピュータに実行せしめることを特徴とするコンピュータプログラム。

【請求項44】 第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、
その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、
その中間画像の生成を、モバイル機器への適用を前提とした機能のもとで行う工程と、
をコンピュータに実行せしめることを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、映像処理技術に関し、とくにデジタル映像効果技術に関する。

【0002】

【従来の技術】BSデジタル放送が開始のタイミングを

迎え、また多数のユーザがPCや形態電話からインターネット上のサービスを楽しむようになり、放送と通信の垣根が急速に崩れつつある。ブロードバンド時代においてマルチメディア文化は真の進展を遂げるであろうし、その中でもとりわけ動画の配信はキーテクノロジーとなる。人間が外界から取得する情報の量において、映像は音声の200倍の内容を伝達できるといわれる。映像は娯楽用途にとどまらず、広く人間の生活と文化を支えるソフトインフラとなる。映像がデジタル化されるにつれ、CGや画像処理技術の適用によって映像効果技術の活躍の場も広がる。

【0003】

【発明が解決しようとする課題】従来より、いろいろな映像効果が提案され、利用されてきた。たとえば、ウォークスルー、多視点画像、特殊フィルタリング、擬似立体視などである。今後、映像の社会における重要性がさらに高まるにつれ、こうした映像効果がより高い品質をもって提供され、同時に伝送系の要請を満たすべく、高いデータ圧縮効果を伴うことが望まれる。この発明はそうした要望に対してなされたものであり、その目的のひとつは、動画の生成を比較的軽いデータで実現する映像効果技術の提供にある。本発明の別の目的は、映像効果に限らず、デジタルおよびインターネット時代における各種画像技術を実現する画像補間技術の提供にある。

【0004】

【課題を解決するための手段】本発明のある態様は、第1画像上に設けられたメッシュの格子点に対応する第2画像上の点を画像マッチングによって検出する工程と、その結果をもとに、第1画像上のメッシュを構成する起点ポリゴンに、第2画像上にて対応する終点ポリゴンを定義する工程とを含む。

【0005】検出する工程は、第1画像と第2画像の間で画素単位のマッチング計算を行ってもよい。この計算は、全画素について行ってもよいし、格子点のみ、または格子点とそれに関連する画素について行ってもよい。

【0006】検出する工程は、第1画像について二次元的な探索を行って検出した特異点と、第2画像について二次元的な探索を行って検出した特異点との対応をもとに画素単位のマッチング計算を行ってもよい。この場合さらに、検出する工程は、第1画像と第2画像を、それぞれ特異点を抽出することによって多重解像度化したうえで、同一解像度レベル間で画像単位のマッチング計算を行い、その結果を異なる解像度レベルにおけるマッチング計算に継承しながら、最終的に最も解像度の細かいレベルにおける画素単位の対応関係を取得してもよい。

【0007】定義する工程は、起点ポリゴンと終点ポリゴンの対応関係を対応点ファイルに出力し、このファイルを第1画像および第2画像と関連づけて保持してもよい。また本発明は、対応点ファイルを利用し、第1画像と第2画像の中間画像をポリゴン単位の補間計算にて生

成する工程をさらに含んでもよい。

【0008】前述の特異点を用いるマッチング方法は、本出願人が先に特許第2927350号にて提案した技術（以下「前提技術」という）の応用であり、前記検出する工程に好適である。ただし、前提技術は格子点またはそれによって決まるポリゴンに関する本発明の特徴にはいっさい触れていない。本発明でポリゴンという一種の簡略化技術を導入したため、前記対応点ファイルのサイズを非常に小さくすることができる。

【0009】すなわち、第1、第2画像がそれぞれ $n \times m$ 画素を有する場合、それらの画素単位の対応をそのまま記述すると、 $(n \times m)^2$ という組合せが生じ、対応点ファイルが膨大なデータ量になる。しかし、これを格子点どうしの対応関係、または実質的に同じことであるが、格子点で決まるポリゴンどうしの対応関係を記述することに改めることでデータ量が大幅に減る。

【0010】一方、第1、第2画像およびこの対応点ファイルがあれば、補間計算で中間画像が生成できる。したがって、本発明は映像効果というモーフィングを実現する。第1、第2画像をともに動画の中のフレームと考えれば、本発明はモーフィングではなく、動画の圧縮技術と把握できる。したがって、第1、第2画像および対応点ファイルのみを保有すれば、動画を再現でき、動画の伝送、ストレージ等において大きな効果を発揮する。

【0011】本発明の別の態様は、映像効果装置に関する。この装置は、第1画像および第2画像を取得する画像入力部と、第1画像と第2画像のマッチングを計算するマッチングプロセッサとを含み、このマッチングプロセッサは、第1画像上のメッシュを構成する起点ポリゴンに、第2画像上にて対応する終点ポリゴンを定義することにより、対応点ファイルを生成する。

【0012】マッチングプロセッサは、第1画像について二次元的な探索を行って検出した特異点と、第2画像について二次元的な探索を行って検出した特異点との対応をもとに画素単位のマッチング計算を行ってもよい。また、第1画像と第2画像を、それぞれ前記特異点を抽出することによって多重解像度化したうえで、同一解像度レベル間で画像単位のマッチング計算を行い、その結果を異なる解像度レベルにおけるマッチング計算に継承しながら、最終的に最も解像度の細かいレベルにおける画素単位の対応関係を取得してもよい。

【0013】この装置はさらに、第1画像、第2画像およびそれらの間の対応点ファイルを取得し、それらふたつの画像の中間画像を生成する中間画像生成部を含み、前記対応点ファイルが第1画像上にとられたメッシュの格子点と、それに対応する第2画像上の点の位置関係を記述していてもよい。また、中間画像、第1画像、第2画像のタイミングを調整してそれらを表示する表示部をさらに含んでもよい。なお、本発明は前提技術を必須とはしない。

【0014】本発明の別のさらに態様は、画像補間方法に関する。この方法は、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得する工程と、その対応情報もとに、第1画像と第2画像の中間画像を生成する工程と、その中間画像の生成に当たり、電子商取引における商品プレゼンテーション機能を提供する工程とを含む。

【0015】その機能に代えて、映像効果に関する機能、動画の生成に関するユーティリティ機能を提供してもよい。また、中間画像の表示をモバイル機器への適用を前提とした機能のもとで行ってもよい。

【0016】商品プレゼンテーション機能は、ユーザからの指示にしたがって商品を回転して表示する機能を含んでもよい。また、商品画像を二方向、例えば左右と上下に補間して得られた中間画像を表示する機能を含んでもよい。

【0017】映像効果に関する機能は、3以上の画像の中から順次ふたつの画像を第1画像および第2画像として選定し中間画像を連続的に表示する機能を含んでもよい。例えば、多数の離散的な視点画像間をつぎつぎ補間して中間画像を生成すれば、あたかもパノラマ的な画像が得られる。視点を時間とともに動かしたかのごとく順に表示すれば、ウォークスルー画像が得られる。映像効果に関する機能は、処理の対象となる画像を二方向に補間して得られた中間画像を表示する機能を含んでもよい。

【0018】動画に関するユーティリティ機能は、生成すべき中間画像の枚数を指定する機能を含んでもよい。多数の中間画像を生成すれば、たとえば、スローモーション画像が生成できる。枚数を指定することで、動画の再生速度をインタラクティブに設定する途がひかれる。生成すべき中間画像の枚数に関する指定はGUIを介してユーザから取得されてもよい。

【0019】モバイル機器への適用を前提とした機能は、第1画像または第2画像の一方のみと前記対応情報をもとに生成された中間画像を表示する機能であってもよい。その場合、画像をテンポラリにしておくためのメモリの容量が少なく済む。対応情報と第1画像のみで中間画像を生成する場合、第1画像の各画素その他の画像要素（以下単に画素で代表させる）を対応情報にしたがって移動させればよい。すなわち、通常は対応情報で画素の位置および画素値すなわち色の両方を補間できるが、これを画素の位置の移動のみの処理に簡素化する。もちろん、対応情報自体に第2画像の画素の色情報を組み込むことも可能であり、その場合は画素の位置と色の両方が良好に補間できる。

【0020】第1画像または第2画像の一方のみと前記対応情報をストーリーミング処理して中間画像が生成されてもよい。この場合も、処理に必要なメモリの容量を減

らす可能性があり、携帯電話その他の端末への実装に有利である。

【0021】本発明のさらに別の態様は、画像を補間する装置であって、第1画像上に設けられたメッシュを構成する起点ポリゴンと第2画像上に設けられたメッシュを構成する終点ポリゴンとの対応情報を取得し、その対応情報もとに第1画像と第2画像の中間画像を生成する中間画像生成部を含み、当該中間画像生成部は、中間画像の生成に当たり、電子商取引における商品プレゼンテーション機能を提供する。この他に、映像効果に関する機能、動画像の生成に関するユーティリティ機能、モバイル機器への適用を前提とした機能を提供してもよい。

【0022】この装置はさらに、中間画像、第1画像、第2画像のタイミングを調整してそれらを表示する表示部を含んでもよい。

【0023】なお、以上の各構成、工程を任意に入れ替えたり、方法と装置の間で表現を一部または全部入れ替え、または追加したり、表現をコンピュータプログラム、記録媒体等に変更したものもまた、本発明として有効である。

【0024】

【発明の実施の形態】はじめに、実施の形態で利用する多重解像度特異点フィルタ技術とそれを用いた画像マッチング処理を「前提技術」として詳述する。これらの技術は本出願人がすでに特許第2927350号を得ている技術であり、本発明との組合せに好適である。本発明では、画像上にメッシュを設け、その格子点によって多数の画素を代表されるため、もともと前提技術のような画素単位のマッチング技術に対する適用効果が高いためである。ただし、実施の形態で採用可能な画像マッチング技術はこれに限られない。

【0025】図18以降、前提技術を利用した画像補間技術および映像効果技術を具体的に説明する。

「前提技術の背景」ふたつの画像の自動的なマッチング、つまり画像領域や画素どうしの対応付けは、コンピュータビジョンやコンピュータグラフィックスにおける最も難しくかつ重要なテーマのひとつである。例えば、あるオブジェクトに関して異なる視点からの画像間でマッチングがとれれば、他の視点からの画像を生成することができる。右目画像と左目画像のマッチングが計算できれば、立体画像を用いた写真測量も可能である。顔の画像のモデルと他の顔の画像のマッチングがとれたとき、目、鼻、口といった特徴的な顔の部分を抽出することができる。例えば人の顔と猫の顔の画像間でマッチングが正確にとれたとき、それらの中割画像を自動的に生成することでモーフィングを完全自動化することができる。

【0026】しかし従来一般に、ふたつの画像間の対応点は人がいちいち指定しなければならず、多大な作業工数を要した。この問題を解消するために数多くの対応点

自動検出方法が提案されている。例えば、エピソード直線を用いることによって対応点の候補の数を減らす考えがある。しかし、その場合でも処理はきわめて複雑である。複雑さを低減するために、左目画像の各点の座標は通常右目画像でもほぼ同じ位置にあると想定される。しかし、こうした制約を設けると、大域的特徴及び局所的特徴を同時に満たすマッチングをとることは非常に困難になる。

【0027】ポリウムレンダリングでは、ボクセルを構成するために一連の断面画像が用いられる。この場合、従来一般に、上方の断面画像における画素が下方の断面画像の同一箇所にある画素と対応すると仮定され、これらの画素のペアが内挿計算に用いられる。このようにきわめて単純な方法を用いるため、連続する断面間の距離が遠く、オブジェクトの断面形状が大きく変化する場合、ポリウムレンダリングで構築されたオブジェクトは不明瞭になりがちである。

【0028】立体写真測量法など、エッジの検出を利用するマッチングアルゴリズムも多い。しかしこの場合、結果的に得られる対応点の数が少ないため、マッチングのとれた対応点間のギャップを埋めるべく、ディスパリティの値を内挿計算しなければならない。一般にあらゆるエッジ検出器は、それらが用いる局所的なウィンドウの中で画素の輝度が変化したとき、これが本当にエッジの存在を示唆するかどうかを判断することが難しい。エッジ検出器は、本来的にすべてハイパスフィルタであり、エッジと同時にノイズも拾ってしまう。

【0029】さらに別の手法として、オプティカルフローが知られている。二枚の画像が与えられたとき、オプティカルフローでは画像内のオブジェクト（剛体）の動きを検出する。その際、オブジェクトの各画素の輝度は変化しないと仮定する。オプティカルフローでは例えば (u, v) のベクトル場の滑らかさといった、いくつかの付加的な条件とともに、各画素の動きベクトル (u, v) を計算する。しかし、オプティカルフローでは画像間の大域的な対応関係を検出することはできない。画素の輝度の局所的な変化に注目するのみであり、画像の変位が大きい場合、システムの誤差は顕著になる。

【0030】画像の大域的な構造を認識するために、多重解像度フィルタも数多く提案されてきた。それらは線形フィルタと非線形フィルタに分類される。前者の例としてウェーブレットがあるが、線形フィルタは一般に、画像マッチングにはさして有用ではない。なぜなら、極値をとる画素の輝度に関する情報がそれらの位置情報とともに次第に不鮮明になるためである。図1(a)と図1(b)は顔の画像に対して平均化フィルタを適用した結果を示している。同図のごとく、極値をとる画素の輝度が平均化によって次第に薄れるとともに、位置も平均化の影響でシフトしていく。その結果、目（輝度の極小点）の輝度や位置の情報は、このような粗い解像度レベ

ルで曖昧になり、この解像度では正しいマッチングを計算することができない。したがって、粗い解像度レベルを設けるのが大域的なマッチングのためでありながら、ここで得られたマッチングは画像の本当の特徴（目、つまり極小点）に正確に対応しない。より精細な解像度レベルで目が鮮明に現れたとしても、大域的なマッチングをとる際に混入した誤差は、もはや取り返しがつかない。入力画像にスムージング処理を加えることにより、テクスチャ領域のステレオ情報が落ちてしまうこともすでに指摘されている。

【0031】一方、最近地形学の分野で利用されはじめた非線形フィルタとして一次元の「ふるい（sieve）」演算子がある。この演算子は、所定の大きさの一次元ウィンドウ内の極小値（または極大値）を選択することにより、縮尺と空間の因果関係を保存しながら画像にスムージング処理を加える。その結果得られる画像は元の画像と同じ大きさであるが、小さな波の成分が取り除かれているため、より単純になる。画像の情報を落とすという点で、この演算子は広い意味での「多重解像度フィルタ」に分類することはできるが、実際にはウェーブレットのように画像の解像度を変えながら画像を階層化するわけではなく（つまり狭い意味での多重解像度フィルタではなく）、画像間の対応の検出には利用できない。

【0032】[前提技術が解決しようとする課題]以上をまとめれば以下の課題が認められる。

1. 画像の特徴を正確に、かつ比較的簡単な処理で把握する画像処理方法が乏しかった。特に、特徴のある点に関する情報、例えば画素値や位置を維持しながら特徴を抽出できる画像処理方法に関する有効な提案が少なかった。
2. 画像の特徴をもとに対応点を自動検出する場合、一般に処理が複雑であるか、ノイズ耐性が低いなどの欠点があった。また、処理に際していろいろな制約を設ける必要があり、大域的特徴及び局所の特徴を同時に満たすマッチングをとることが困難だった。
3. 画像の大域的な構造または特徴を認識するために多重解像度フィルタを導入しても、そのフィルタが線形フィルタの場合、画素の輝度情報と位置情報が曖昧になった。その結果、対応点の把握が不正確になりやすかった。非線形フィルタである一次元ふるい演算子は画像を階層化しないため、画像間の対応点の検出には利用できなかった。
4. これらの結果、対応点を正しく把握しようとすれば、結局人手による指定に頼るほか有効な手だてがなかった。

【0033】前提技術はこれらの課題の解決を目的としてなされたものであり、画像処理の分野において、画像の特徴の的確な把握を可能にする技術を提供するものである。

【0034】[前提技術が課題を解決するための手段]こ

の目的のために前提技術のある態様は、新たな多重解像度の画像フィルタを提案する。この多重解像度フィルタは画像から特異点を抽出する。したがって、特異点フィルタともよばれる。特異点とは画像上特徴をもつ点をいう。例として、ある領域において画素値（画素値とは、色番号、輝度値など画像または画素に関する任意の数値を指す）が最大になる極大点、最小になる極小点、ある方向については最大だが別の方向については最小になるような鞍点がある。特異点は位相幾何学上の概念であってもよい。ただし、その他どのような特徴を有してもよい。いかなる性質の点を特異点と考えるかは、前提技術にとって本質問題ではない。

【0035】この態様では、多重解像度フィルタを用いた画像処理が行われる。まず検出工程において、第一の画像に対し、二次元的な探索を行って特異点が検出される。つぎに生成工程において、検出された特異点を抽出して第一の画像よりも解像度の低い第二の画像が生成される。第二の画像には第一の画像のもつ特異点が引き継がれる。第二の画像は第一の画像よりも解像度が低いため、画像の大域的な特徴の把握に好適である。

【0036】前提技術の別の態様は特異点フィルタを用いた画像マッチング方法に関する。この態様では、始点画像と終点画像間のマッチングがとられる。始点画像および終点画像とは、ふたつの画像の区別のために便宜的に与えた名称であり、本質的な違いはない。

【0037】この態様では、第一工程にて、始点画像に特異点フィルタを施して解像度の異なる一連の始点階層画像が生成される。第二工程では、終点画像に特異点フィルタを施して解像度の異なる一連の終点階層画像が生成される。始点階層画像、終点階層画像とは、それぞれ始点画像、終点画像を階層化して得られる画像群をいい、それぞれ最低2枚の画像からなる。つぎに第三工程において、始点階層画像と終点階層画像のマッチングが解像度レベルの階層の中で計算される。この態様によれば、多重解像度フィルタによって特異点に関連する画像の特徴が抽出され、および／または明確化されるため、マッチングが容易になる。マッチングのための拘束条件は特に必要としない。

【0038】前提技術のさらに別の態様も始点画像と終点画像のマッチングに関する。この態様では、予め複数のマッチング評価項目のそれぞれに関して評価式を設け、それらの評価式を統合して総合評価式を定義し、その総合評価式の極値付近に注目して最適マッチングを探索する。総合評価式は、評価式の少なくともひとつに係数パラメータを掛けたうえでそれらの評価式の総和として定義してもよく、その場合、総合評価式またはいずれかの評価式がほぼ極値をとる状態を検出して前記パラメータを決定してもよい。「極値付近」または「ほぼ極値をとる」としたのは、多少誤差を含んでいてもよいためである。多少の誤差は前提技術にはさして問題とならな

い。

【0039】極値自体も前記パラメータに依存するため、極値の挙動、つまり極値の変化の様子をもとに、最適と考えられるパラメータを決定する余地が生じる。この態様はその事実を利用している。この態様によれば、元来調整の困難なパラメータの決定を自動化する途が拓かれる。

【0040】[前提技術の実施の形態]最初に[1]で前提技術の要素技術の詳述し、[2]で処理手順を具体的に説明する。さらに[3]で実験の結果を報告する。

【0041】[1] 要素技術の詳細

[1. 1] イントロダクション

特異点フィルタと呼ばれる新たな多重解像度フィルタを導入し、画像間のマッチングを正確に計算する。オブジェクトに関する予備知識は一切不要である。画像間のマッチングの計算は、解像度の階層を進む間、各解像度において計算される。その際、粗いレベルから精細なレベルへと順に解像度の階層を辿っていく。計算に必要なパラメータは、人間の視覚システムに似た動的計算によって完全に自動設定される。画像間の対応点を人手で特定する必要はない。

【0042】本前提技術は、例えば完全に自動的なモーフィング、物体認識、立体写真測量、ボリュームレンダリング、少ないフレームからの滑らかな動画の生成な

どに応用できる。モーフィングに用いる場合、与えられた画像を自動的に変形することができる。ボリュームレンダリングに用いる場合、断面間の中間的な画像を正確に再構築することができる。断面間の距離が遠く、断面の形状が大きく変化する場合でも同様である。

【0043】[1. 2] 特異点フィルタの階層

前提技術に係る多重解像度特異点フィルタは、画像の解像度を落としながら、しかも画像に含まれる各特異点の輝度及び位置を保存することができる。ここで画像の幅をN、高さをMとする。以下簡単のため、 $N=M=2^n$ (n は自然数)と仮定する。また、区間 $[0, N] \subset \mathbb{R}$ を I と記述する。 (i, j) における画像の画素を $p(i, j)$ と記述する ($i, j \in I$)。

【0044】ここで多重解像度の階層を導入する。階層化された画像群は多重解像度フィルタで生成される。多重解像度フィルタは、もとの画像に対して二次元的な探索を行って特異点を検出し、検出された特異点を抽出してもとの画像よりも解像度の低い別の画像を生成する。ここで第 m レベルにおける各画像のサイズは $2^m \times 2^m$ ($0 \leq m \leq n$) とする。特異点フィルタは次の4種類の新たな階層画像を n から下がる方向で再帰的に構築する。

【0045】

【数1】

$$\begin{aligned} p_{(i,j)}^{(m,0)} &= \min(\min(p_{(2i,2j)}^{(m+1,0)}, p_{(2i,2j+1)}^{(m+1,0)}), \min(p_{(2i+1,2j)}^{(m+1,0)}, p_{(2i+1,2j+1)}^{(m+1,0)})) \\ p_{(i,j)}^{(m,1)} &= \max(\min(p_{(2i,2j)}^{(m+1,1)}, p_{(2i,2j+1)}^{(m+1,1)}), \min(p_{(2i+1,2j)}^{(m+1,1)}, p_{(2i+1,2j+1)}^{(m+1,1)})) \\ p_{(i,j)}^{(m,2)} &= \min(\max(p_{(2i,2j)}^{(m+1,2)}, p_{(2i,2j+1)}^{(m+1,2)}), \max(p_{(2i+1,2j)}^{(m+1,2)}, p_{(2i+1,2j+1)}^{(m+1,2)})) \\ p_{(i,j)}^{(m,3)} &= \max(\max(p_{(2i,2j)}^{(m+1,3)}, p_{(2i,2j+1)}^{(m+1,3)}), \max(p_{(2i+1,2j)}^{(m+1,3)}, p_{(2i+1,2j+1)}^{(m+1,3)})) \end{aligned}$$

(式1)

ただしここで、

【数2】

$$p_{(i,j)}^{(n,0)} = p_{(i,j)}^{(n,1)} = p_{(i,j)}^{(n,2)} = p_{(i,j)}^{(n,3)} = p_{(i,j)} \quad (\text{式2})$$

とする。以降これら4つの画像を副画像(サブイメージ)と呼ぶ。 $\min_{x \leq t \leq x+1}$ 、 $\max_{x \leq t \leq x+1}$ をそれぞれ α 及び β と記述すると、副画像はそれぞれ以下のように記述できる。

【0046】

$$P(m, 0) = \alpha(x) \alpha(y) p(m+1, 0)$$

$$P(m, 1) = \alpha(x) \beta(y) p(m+1, 1)$$

$$P(m, 2) = \beta(x) \alpha(y) p(m+1, 2)$$

$$P(m, 3) = \beta(x) \beta(y) p(m+1, 3)$$

すなわち、これらは α と β のテンソル積のようなものと考えられる。副画像はそれぞれ特異点に対応している。これらの式から明らかなように、特異点フィルタはもと

の画像について 2×2 画素で構成されるブロックごとに特異点を検出する。その際、各ブロックのふたつの方向、つまり縦と横について、最大画素値または最小画素値をもつ点を探索する。画素値として、前提技術では輝度を採用するが、画像に関するいろいろな数値を採用することができる。ふたつの方向の両方について最大画素値となる画素は極大点、ふたつの方向の両方について最小画素値となる画素は極小点、ふたつの方向の一方について最大画素値となるとともに、他方について最小画素値となる画素は鞍点として検出される。

【0047】特異点フィルタは、各ブロックの内部で検出された特異点の画像(ここでは1画素)でそのブロッ

クの画像（ここでは4画素）を代表させることにより、画像の解像度を落とす。特異点の理論的な観点からすれば、 $\alpha(x)\alpha(y)$ は極小点を保存し、 $\beta(x)\beta(y)$ は極大点を保存し、 $\alpha(x)\beta(y)$ 及び $\beta(x)\alpha(y)$ は鞍点を保存する。

【0048】はじめに、マッチングをとるべき始点（ソース）画像と終点（デスティネーション）画像に対して別々に特異点フィルタ処理を施し、それぞれ一連の画像群、すなわち始点階層画像と終点階層画像を生成しておく。始点階層画像と終点階層画像は、特異点の種類に対応してそれぞれ4種類ずつ生成される。

【0049】この後、一連の解像度レベルの中で始点階層画像と終点階層画像のマッチングがとれられていく。まず $p(m, 0)$ を用いて極小点のマッチングがとられる。次に、その結果に基づき、 $p(m, 1)$ を用いて鞍点のマッチングがとられ、 $p(m, 2)$ を用いて他の鞍点のマッチングがとられる。そして最後に $p(m, 3)$ を用いて極大点のマッチングがとられる。

【0050】図1(c)と図1(d)はそれぞれ図1(a)と図1(b)の副画像 $p(5, 0)$ を示している。同様に、図1(e)と図1(f)は $p(5, 1)$ 、図1(g)と図1(h)は $p(5, 2)$ 、図1(i)と図1(j)は $p(5, 3)$ をそれぞれ示している。これらの図からわかるとおり、副画像によれば画像の特徴部分のマッチングが容易になる。まず $p(5, 0)$ によって目が明確になる。目は顔の中で輝度の極小点だからである。 $p(5, 1)$ によれば口が明確になる。口は横方向で輝度が低いためである。 $p(5, 2)$ によれば首の

両側の縦線が明確になる。最後に、 $p(5, 3)$ によって耳や頬の最も明るい点が明確になる。これらは輝度の極大点だからである。

【0051】特異点フィルタによれば画像の特徴が抽出できるため、例えばカメラで撮影された画像の特徴と、予め記録しておいたいくつかのオブジェクトの特徴を比較することにより、カメラに映った被写体を識別することができる。

【0052】[1. 3] 画像間の写像の計算

始点画像の位置 (i, j) の画素を $p^{(n)}(i, j)$ と書き、同じく終点画像の位置 (k, l) の画素を $q^{(n)}(k, l)$ で記述する。 $i, j, k, l \in I$ とする。画像間の写像のエネルギー（後述）を定義する。このエネルギーは、始点画像の画素の輝度と終点画像の対応する画素の輝度の差、及び写像の滑らかさによって決まる。最初に最小のエネルギーを持つ $p(m, 0)$ と $q(m, 0)$ 間の写像 $f(m, 0) : p(m, 0) \rightarrow q(m, 0)$ が計算される。 $f(m, 0)$ に基づき、最小エネルギーを持つ $p(m, 1)$ 、 $q(m, 1)$ 間の写像 $f(m, 1)$ が計算される。この手続は、 $p(m, 3)$ と $q(m, 3)$ の間の写像 $f(m, 3)$ の計算が終了するまで続く。各写像 $f(m, i)$ ($i=0, 1, 2, \dots$)を副写像と呼ぶことにする。 $f(m, i)$ の計算の都合のために、 i の順序は次式のように並べ替えることができる。並べ替えが必要な理由は後述する。

【0053】

【数3】

$$f(m, i) : p^{(m, \sigma(i))} \rightarrow q^{(m, \sigma(i))}$$

(式3)

ここで $\sigma(i) \in \{0, 1, 2, 3\}$ である。

【0054】[1. 3. 1] 全単射

始点画像と終点画像の間のマッチングを写像で表現する場合、その写像は両画像間で全単射条件を満たすべきである。両画像に概念上の優劣はなく、互いの画素が全射かつ単射で接続されるべきだからである。しかしながら通常の場合とは異なり、ここで構築すべき写像は全単射のデジタル版である。前提技術では、画素は格子点によって特定される。

【0055】始点副画像（始点画像について設けられた副画像）から終点副画像（終点画像について設けられた副画像）への写像は、 $f(m, s) : I/2^{n-m} \times I/2^{n-m} \rightarrow I/2^{n-m} \times I/2^{n-m}$ ($s=0,$

$$p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)}$$

(式4)

を考える ($i=0, \dots, 2^{n-m}-1, j=0, \dots, 2^{n-m}-1$)。ここで R の各辺（エッジ）の方向を以下のように定める。

1, ...)によって表される。ここで、 $f(m, s)(i, j) = (k, l)$ は、始点画像の $p(m, s)(i, j)$ が終点画像の $q(m, s)(k, l)$ に写像されることを意味する。簡単のために、 $f(i, j) = (k, l)$ が成り立つとき画素 $q(k, l)$ を $q_f(i, j)$ と記述する。

【0056】前提技術で扱う画素（格子点）のようにデータが離散的な場合、全単射の定義は重要である。ここでは以下のように定義する (i, i', j, j', k, l は全て整数とする)。まず始めに、始点画像の平面において R によって表記される各正方形領域、

【数4】

【0057】

【数5】

$$\overrightarrow{p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)}}, \overrightarrow{p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)}}, \overrightarrow{p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)}} \text{ and } \overrightarrow{p_{(i,j+1)}^{(m,s)} p_{(i,j)}^{(m,s)}} \quad (\text{式5})$$

この正方形は写像 f によって終点画像平面における四辺形に写像されなければならない。 $f(m, s)(R)$ に

よって示される四辺形、

【数6】

$$q_{f(i,j)}^{(m,s)} q_{f(i+1,j)}^{(m,s)} q_{f(i+1,j+1)}^{(m,s)} q_{f(i,j+1)}^{(m,s)} \quad (\text{式6})$$

は、以下の全単射条件を満たす必要がある。

【0058】1. 四辺形 $f(m, s)(R)$ のエッジは互いに交差しない。

2. $f(m, s)(R)$ のエッジの方向は R のそれらに等しい(図2の場合、時計回り)。

3. 緩和条件として収縮写像(リトラクション: retractions)を許す。

【0059】何らかの緩和条件を設けないかぎり、全単射条件を完全に満たす写像は単位写像しかないためである。ここでは $f(m, s)(R)$ のひとつのエッジの長さが0、すなわち $f(m, s)(R)$ は三角形になってもよい。しかし、面積が0となるような図形、すなわち1点または1本の線分になってはならない。図2(R)がもとの四辺形の場合、図2(A)と図2(D)は全単射条件を満たすが、図2(B)、図2(C)、図2

(E)は満たさない。

【0060】実際のインプリメンテーションでは、写像

$$C_{(i,j)}^{(m,s)} = |V(p_{(i,j)}^{(m,s)}) - V(q_{f(i,j)}^{(m,s)})|^2 \quad (\text{式7})$$

ここで、 $V(p_{(i,j)}^{(m,s)})$ 及び $V(q_{f(i,j)}^{(m,s)})$ はそれぞれ画素 $p_{(i,j)}^{(m,s)}$ 及び $q_{f(i,j)}^{(m,s)}$ の輝度である。 f のトータルエネルギー $C(m, s)$ は、マッチ

$$C_f^{(m,s)} = \sum_{i=0}^{i=2^m-1} \sum_{j=0}^{j=2^m-1} C_{(i,j)}^{(m,s)} \quad (\text{式8})$$

[1. 3. 2. 2] 滑らかな写像のための画素の位置に関するコスト

滑らかな写像を得るために、写像に関する別のエネルギー D_f を導入する。このエネルギーは画素の輝度とは関係なく、 $p_{(i,j)}^{(m,s)}$ および $q_{f(i,j)}^{(m,s)}$ の位置によって決まる ($i=0, \dots, 2^m$

$$D_{(i,j)}^{(m,s)} = \eta E_{0(i,j)}^{(m,s)} + E_{1(i,j)}^{(m,s)} \quad (\text{式9})$$

ただし、係数パラメータ η は0以上の実数であり、また、

$$E_{0(i,j)}^{(m,s)} = \|(i, j) - f^{(m,s)}(i, j)\|^2 \quad (\text{式10})$$

【数11】

が全射であることを容易に保証すべく、さらに以下の条件を課してもよい。つまり始点画像の境界上の各画素は、終点画像において同じ位置を占める画素に写影されるというものである。すなわち、 $f(i, j) = (i, j)$ (ただし $i=0, i=2^m-1, j=0, j=2^m-1$ の4本の線上)である。この条件を以下「付加条件」とも呼ぶ。

【0061】[1. 3. 2] 写像のエネルギー

[1. 3. 2. 1] 画素の輝度に関するコスト

写像 f のエネルギーを定義する。エネルギーが最小になる写像を探すことが目的である。エネルギーは主に、始点画像の画素の輝度とそれに対応する終点画像の画素の輝度の差で決まる。すなわち、写像 $f(m, s)$ の点 (i, j) におけるエネルギー $C(m, s)(i, j)$ は次式によって定まる。

【0062】

【数7】

ングを評価するひとつの評価式であり、つぎに示す $C(m, s)(i, j)$ の合計で定義できる。

【0063】

【数8】

$-1, j=0, \dots, 2^m-1$)。点 (i, j) における写像 $f(m, s)$ のエネルギー $D(m, s)(i, j)$ は次式で定義される。

【0064】

【数9】

【数10】

$$E_{1(i,j)}^{(m,s)} = \sum_{i'=i-1}^i \sum_{j'=j-1}^j \|(f^{(m,s)}(i,j) - (i,j)) - (f^{(m,s)}(i',j') - (i',j'))\|^2/4$$

(式11)

とする。ここで、

$$\|(x,y)\| = \sqrt{x^2 + y^2}$$

であり、 $i' < 0$ および $j' < 0$ に対して $f(i', j')$ は0と決める。 E_0 は (i, j) 及び $f(i, j)$ の距離で決まる。 E_0 は画素があまりにも離れた画素へ写影されることを防ぐ。ただし E_0 は、後に別のエネルギー関数で置き換える。 E_1 は写像の滑らかさを保証する。 E_1 は、 $p(i, j)$ の変位とその隣接点の変

【数12】

(式12)

位の間の隔たりを表す。以上の考察をもとに、マッチングを評価する別の評価式であるエネルギー D_f は次式で定まる。

【0065】

【数13】

$$D_f^{(m,s)} = \sum_{i=0}^{i=2^m-1} \sum_{j=0}^{j=2^m-1} D_{(i,j)}^{(m,s)}$$

(式13)

[1.3.2.3] 写像の総エネルギー

写像の総エネルギー、すなわち複数の評価式の統合に係る総合評価式は $\lambda C^{(m,s)}_f + D^{(m,s)}_f$ で定義される。ここで係数パラメータ λ は0以上の実数である。目的は総合評価式が極値をとる状態を検出するこ

と、すなわち次式で示す最小エネルギーを与える写像を見いだすことである。

【0066】

【数14】

$$\min_f \lambda C_f^{(m,s)} + D_f^{(m,s)}$$

(式14)

$\lambda = 0$ 及び $\eta = 0$ の場合、写像は単位写像になることに注意すべきである（すなわち、全ての $i = 0, \dots, 2^m - 1$ 及び $j = 0, \dots, 2^m - 1$ に対して f

$(m, s)(i, j) = (i, j)$ となる）。後述のごとく、本前提技術では最初に $\lambda = 0$ 及び $\eta = 0$ の場合を評価するため、写像を単位写像から徐々に変形していくことができる。仮に総合評価式の λ の位置を変えて $C^{(m,s)}_f + \lambda D^{(m,s)}_f$ と定義したとすれば、 $\lambda = 0$ 及び $\eta = 0$ の場合に総合評価式が $C^{(m,s)}_f$ だけになり、本来何等関連のない画素どうしが単に輝度が近いというだけで対応づけられ、写像が無意味なものになる。そうした無意味な写像をもとに写像を変形していてもまったく意味をなさない。このため、単位写像が評価の開始時点で最良の写像として選択されるよう係数パラメータの与えかたが配慮されている。

【0067】オプティカルフローもこの前提技術同様、画素の輝度の差と滑らかさを考慮する。しかし、オプティカルフローは画像の変換に用いることはできない。オ

$$(i', j') = (\lfloor \frac{i}{2} \rfloor, \lfloor \frac{j}{2} \rfloor)$$

が成り立つとき、 $p^{(m-1,s)}(i', j')$ 、 $q^{(m-1,s)}(i', j')$ をそれぞれ $p^{(m,s)}(i, j)$ 、 $q^{(m,s)}(i, j)$ の parent と呼ぶことにする。 $[x]$ は x を越えない最大整数であ

ブジェクトの局所的な動きしか考慮しないためである。前提技術に係る特異点フィルタを用いることによって大域的な対応関係を検出することができる。

【0068】[1.3.3] 多重解像度の導入による写像の決定

最小エネルギーを与え、全単射条件を満足する写像 f_{min} を多重解像度の階層を用いて求める。各解像度レベルにおいて始点副画像及び終点副画像間の写像を計算する。解像度の階層の最上位（最も粗いレベル）からスタートし、各解像度レベルの写像を、他のレベルの写像を考慮に入れながら決定する。各レベルにおける写像の候補の数は、より高い、つまりより粗いレベルの写像を用いることによって制限される。より具体的には、あるレベルにおける写像の決定に際し、それよりひとつ粗いレベルにおいて求められた写像が一種の拘束条件として課される。

【0069】まず、

【数15】

(式15)

る。また $p^{(m,s)}(i, j)$ 、 $q^{(m,s)}(i, j)$ をそれぞれ $p^{(m-1,s)}(i', j')$ 、 $q^{(m-1,s)}(i', j')$ の child と呼ぶ。関数 parent

(i, j) は次式で定義される。

【0070】

【数16】

$$\text{parent}(i, j) = \left(\left\lfloor \frac{i}{2} \right\rfloor, \left\lfloor \frac{j}{2} \right\rfloor \right) \quad (\text{式16})$$

$p(m, s)(i, j)$ と $q(m, s)(k, l)$ の間の写像 $f(m, s)$ は、エネルギー計算を行って最小になったものを見つけることで決定される。 $f(m, s)(i, j) = (k, l)$ の値は $f(m-1, s)(m=1, 2, \dots, n)$ を用いることによって、以下のように

決定される。まず、 $q(m, s)(k, l)$ は次の四辺形の内部になければならないという条件を課し、全単射条件を満たす写像のうち現実性の高いものを絞り込む。

【0071】

【数17】

$$q_{g(m,s)}^{(m,s)}(i-1,j-1) q_{g(m,s)}^{(m,s)}(i-1,j+1) q_{g(m,s)}^{(m,s)}(i+1,j+1) q_{g(m,s)}^{(m,s)}(i+1,j-1)$$

(式17)

ただしここで、

【数18】

$$g^{(m,s)}(i, j) = f^{(m-1,s)}(\text{parent}(i, j)) + f^{(m-1,s)}(\text{parent}(i, j) + (1, 1))$$

(式18)

である。こうして定めた四辺形を、以下 $p(m, s)(i, j)$ の相続(inherited)四辺形と呼ぶことにする。相続四辺形の内部において、エネルギーを最小にする画素を求める。

【0072】図3は以上の手順を示している。同図において、始点画像のA, B, C, Dの画素は、第 $m-1$ レベルにおいてそれぞれ終点画像のA', B', C', D'へ写影される。画素 $p(m, s)(i, j)$ は、相続四辺形A' B' C' D'の内部に存在する画素 q

$(m, s) f(m)(i, j)$ へ写影されなければならない。以上の配慮により、第 $m-1$ レベルの写像から第 m レベルの写像への橋渡しがなされる。

【0073】先に定義したエネルギー E_0 は、第 m レベルにおける副写像 $f(m, 0)$ を計算するために、次式に置き換える。

【0074】

【数19】

$$E_{0(i,j)} = \|f^{(m,0)}(i, j) - g^{(m)}(i, j)\|^2 \quad (\text{式19})$$

また、副写像 $f(m, s)$ を計算するためには次式を用いる。

【0075】

【数20】

$$E_{0(i,j)} = \|f^{(m,s)}(i, j) - f^{(m,s-1)}(i, j)\|^2 \quad (1 \leq i) \quad (\text{式20})$$

こうしてすべての副写像のエネルギーを低い値に保つ写像が得られる。式20により、異なる特異点に対応する副写像が、副写像どうしの類似度が高くなるように同一レベル内で関連づけられる。式19は、 f

$(m, s)(i, j)$ と、第 $m-1$ レベルの画素の一部と考えた場合の (i, j) が射影されるべき点の位置との距離を示している。

【0076】仮に、相続四辺形A' B' C' D'の内部に全単射条件を満たす画素が存在しない場合は以下の措置をとる。まず、A' B' C' D'の境界線からの距離がL(始めはL=1)である画素を調べる。それらのうち、エネルギーが最小になるものが全単射条件を満たせば、これを $f(m, s)(i, j)$ の値として選択する。そのような点が発見されるか、またはLがその上限の $L(m) \max$ に到達するまで、Lを大きくしてい

く。 $L(m) \max$ は各レベル m に対して固定である。そのような点が発見されない場合、全単射の第3の条件を一時的に無視して変換先の四辺形の面積がゼロになるような写像も認め、 $f(m, s)(i, j)$ を決定する。それでも条件を満たす点が見つからない場合、つぎに全単射の第1及び第2条件を外す。

【0077】多重解像度を用いる近似法は、写像が画像の細部に影響されることを回避しつつ、画像間の大域的な対応関係を決定するために必須である。多重解像度による近似法を用いなければ、距離の遠い画素間の対応関係を見いだすことは不可能である。その場合、画像のサイズはきわめて小さなものに限定しなければならず、変化の小さな画像しか扱うことができない。さらに、通常写像に滑らかさを要求するため、そうした画素間の対応関係を見つけにくくしている。距離のある画素から画素

への写像のエネルギーは高いためである。多重解像度を用いた近似法によれば、そうした画素間の適切な対応関係を見いだすことができる。それらの距離は、解像度の階層の上位レベル（粗いレベル）において小さいためである。

【0078】[1. 4] 最適なパラメータ値の自動決定
既存のマッチング技術の主な欠点のひとつに、パラメータ調整の困難さがある。大抵の場合、パラメータの調整は人手作業によって行われ、最適な値を選択することはきわめて難しい。前提技術に係る方法によれば、最適なパラメータ値を完全に自動決定することができる。

【0079】前提技術に係るシステムはふたつのパラメータ、 λ 及び η を含む。端的に言えば、 λ は画素の輝度の差の重みであり、 η は写像の剛性を示している。これらのパラメータの値は初期値が0であり、まず $\eta = 0$ に固定して λ を0から徐々に増加させる。 λ の値を大きくしながら、しかも総合評価式（式14）の値を最小にする場合、各副写像に関する $C(m, s)_f$ の値は一般に小さくなっていく。このことは基本的にふたつの画像がよりマッチしなければならないことを意味する。しかし、 λ が最適値を超えると以下の現象が発生する。

【0080】1. 本来対応すべきではない画素どうしが、単に輝度が近いというだけで誤って対応づけられる。

2. その結果、画素どうしの対応関係がおかしくなり、写像がくずれはじめる。

【0081】3. その結果、式14において $D(m, s)_f$ が急激に増加しようとする。

4. その結果、式14の値が急激に増加しようとするため、 $D(m, s)_f$ の急激な増加を抑制するよう $f(m, s)$ が変化し、その結果 $C(m, s)_f$ が増加する。したがって、 λ を増加させながら式14が最小値を

とるという状態を維持しつつ $C(m, s)_f$ が減少から増加に転じる閾値を検出し、その λ を $\eta = 0$ における最適値とする。つぎに η を少しづつ増やして $C(m, s)_f$ の挙動を検査し、後述の方法で η を自動決定する。その η に対応して λ も決まる。

【0082】この方法は、人間の視覚システムの焦点機構の動作に似ている。人間の視覚システムでは、一方の目を動かしながら左右両目の画像のマッチングがとられる。オブジェクトがはっきりと認識できるとき、その目が固定される。

【0083】[1. 4. 1] λ の動的決定

λ は0から所定の刻み幅で増加されていき、 λ の値が変わる度に副写像が評価される。式14のごとく、総エネルギーは $\lambda C(m, s)_f + D(m, s)_f$ によって定義される。式9の $D(m, s)_f$ は滑らかさを表すもので、理論的には単位写像の場合に最小になり、写像が歪むほど E_0 も E_1 も増加していく。 E_1 は整数であるから、 $D(m, s)_f$ の最小刻み幅は1である。このため、現在の $\lambda C(m, s)_{(i, j)}$ の変化（減少量）が1以上でなければ、写像を変化させることによって総エネルギーを減らすことはできない。なぜなら、写像の変化に伴って $D(m, s)_f$ は1以上増加するため、 $\lambda C(m, s)_{(i, j)}$ が1以上減少しない限り総エネルギーは減らないためである。

【0084】この条件のもと、 λ の増加に伴い、正常な場合に $C(m, s)_{(i, j)}$ が減少することを示す。 $C(m, s)_{(i, j)}$ のヒストグラムを $h(l)$ と記述する。 $h(l)$ はエネルギー $C(m, s)_{(i, j)}$ が l^2 である画素の数である。 $\lambda l^2 \geq 1$ が成り立つために、例えば $l^2 = 1/\lambda$ の場合を考える。 λ が λ_1 から λ_2 まで微小量変化するとき、

【数21】

$$A = \sum_{l=\lceil \frac{1}{\lambda_2} \rceil}^{\lfloor \frac{1}{\lambda_1} \rfloor} h(l) \simeq \int_{l=\frac{1}{\lambda_2}}^{\frac{1}{\lambda_1}} h(l) dl = - \int_{\lambda_2}^{\lambda_1} h(l) \frac{1}{\lambda^{3/2}} d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{h(l)}{\lambda^{3/2}} d\lambda$$

(式21)

で示される A 個の画素が、

$$C_f^{(m, s)} - l^2 = C_f^{(m, s)} - \frac{1}{\lambda}$$

のエネルギーを持つより安定的な状態に変化する。ここでは仮に、これらの画素のエネルギーがすべてゼロにな

$$\partial C_f^{(m, s)} = -\frac{A}{\lambda}$$

だけ変化することを示し、その結果、

$$\frac{\partial C_f^{(m, s)}}{\partial \lambda} = -\frac{h(l)}{\lambda^{5/2}}$$

が成立する。 $h(l) > 0$ であるから、通常 C

【数22】

(式22)

と近似している。この式は $C(m, s)_f$ の値が、

【数23】

(式23)

【数24】

(式24)

$C(m, s)_f$ は減少する。しかし、 λ が最適値を越えよ

うとするとき、上述の現象、つまり $C(m, s)_f$ の増加が発生する。この現象を検出することにより、 λ の最適値を決定する。

$$h(l) = Hl^k = \frac{H}{\lambda^{k/2}} \quad (\text{式25})$$

と仮定すれば、

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{H}{\lambda^{5/2+k/2}} \quad (\text{式26})$$

が成り立つ。このとき $k \neq -3$ であれば、

$$C_f^{(m,s)} = C + \frac{H}{(3/2 + k/2)\lambda^{3/2+k/2}} \quad (\text{式27})$$

となる。これが $C(m, s)_f$ の一般式である (C は定数)。

【0086】 λ の最適値を検出する際、さらに安全を見て、全単射条件を破る画素の数を検査してもよい。ここで各画素の写像を決定する際、全単射条件を破る確率を p_0 と仮定する。この場合、

【数28】

$$\frac{\partial A}{\partial \lambda} = \frac{h(l)}{\lambda^{3/2}} \quad (\text{式28})$$

が成立しているため、全単射条件を破る画素の数は次式の率で増加する。

【0087】

【数29】

$$B_0 = \frac{h(l)p_0}{\lambda^{3/2}} \quad (\text{式29})$$

従って、

【数30】

$$\frac{B_0 \lambda^{3/2}}{p_0 h(l)} = 1 \quad (\text{式30})$$

は定数である。仮に $h(1) = H1^k$ を仮定するとき、例えば、

【数31】

$$B_0 \lambda^{3/2+k/2} = p_0 H \quad (\text{式31})$$

は定数になる。しかし λ が最適値を越えると、上の値は急速に増加する。この現象を検出し、 $B_0 \lambda^{3/2+k/2}$ の値が異常値 B_{0thres} を越えるかどうかを検査し、 λ の最適値を決定することがで

【0085】なお、 $H(h > 0)$ 及び k を定数とするとき、

【数25】

【数26】

【数27】

きる。同様に、 $B_1 \lambda^{3/2+k/2}$ の値が異常値 B_{1thres} を越えるかどうかを検査することにより、全単射の第3の条件を破る画素の増加率 B_1 を確認する。ファクター $2m$ を導入する理由は後述する。このシステムはこれら2つの閾値に敏感ではない。これらの閾値は、エネルギー $C(m, s)_f$ の観察では検出し損なった写像の過度の歪みを検出するために用いることができる。

【0088】なお実験では、副写像 $f(m, s)$ を計算する際、もし λ が 0.1 を越えたら $f(m, s)$ の計算は止めて $f(m, s+1)$ の計算に移行した。 $\lambda > 0.1$ のとき、画素の輝度 255 レベル中のわずかに「3」の違いが副写像の計算に影響したためであり、 $\lambda > 0.1$ のとき正しい結果を得ることは困難だったためである。

【0089】[1. 4. 2] ヒストグラム $h(1)$ $C(m, s)_f$ の検査はヒストグラム $h(1)$ に依存しない。全単射及びその第3の条件の検査の際、 $h(1)$ に影響を受けうる。実際に $(\lambda, C(m, s)_f)$ をプロットすると、 k は通常1付近にある。実験では $k=1$ を用い、 $B_0 \lambda^2$ と $B_1 \lambda^2$ を検査した。仮に k の本当の値が1未満であれば、 $B_0 \lambda^2$ と $B_1 \lambda^2$ は定数にならず、ファクター $\lambda^{(1-k)/2}$ に従って徐々に増加する。 $h(1)$ が定数であれば、例えばファクターは $\lambda^{1/2}$ である。しかし、こうした差は閾値 B_{0thres} を正しく設定することによって吸収することができる。

【0090】ここで次式のごとく始点画像を中心が (x_0, y_0) 、半径 r の円形のオブジェクトであると仮定する。

【数32】

$$p(i,j) = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_0)^2 + (j-y_0)^2}) & (\sqrt{(i-x_0)^2 + (j-y_0)^2} \leq r) \\ 0 & (\text{otherwise}) \end{cases} \quad (\text{式32})$$

一方、終点画像は、次式のごとく中心 (x_1, y_1) 、

半径が r のオブジェクトであるとする。

【数33】

$$q_{(i,j)} = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_1)^2 + (j-y_1)^2}) & (\sqrt{(i-x_1)^2 + (j-y_1)^2} \leq r) \\ 0 & (\text{otherwise}) \end{cases} \quad (\text{式33})$$

ここで $c(x)$ は $c(x) = x^k$ の形であるとする。中心 (x_0, y_0) 及び (x_1, y_1) が十分遠い場合、ヒストグラム $h(1)$ は次式の形となる。

【数34】

$$h(l) \propto r l^k \quad (k \neq 0)$$

(式34)

$k=1$ のとき、画像は背景に埋め込まれた鮮明な境界線を持つオブジェクトを示す。このオブジェクトは中心が暗く、周囲にいくに従って明るくなる。 $k=-1$ のとき、画像は曖昧な境界線を持つオブジェクトを表す。このオブジェクトは中心が最も明るく、周囲にいくに従って暗くなる。一般のオブジェクトはこれらふたつのタイプのオブジェクトの中間にあると考えてもさして一般性を失わない。したがって、 k は $-1 \leq k \leq 1$ として大抵の場合をカバーでき、式27が一般に減少関数であるこ

$$E_{0(i,j)}^{(m,s)} = \|f^{(m,s)}(i,j) - f^{(m,s-1)}(i,j)\|^2$$

η が0のとき、 $D^{(n)}_f$ は直前の副写像と無関係に決定され、現在の副写像は弾性的に変形され、過度に歪むことになる。一方、 η が非常に大きな値のとき、 $D^{(n)}_f$ は直前の副写像によってほぼ完全に決まる。このとき副写像は非常に剛性が高く、画素は同じ場所に射影される。その結果、写像は単位写像になる。 η の値が0から次第に増えるとき、後述のごとく $C^{(n)}_f$ は徐々に減少する。しかし η の値が最適値を越えると、図4に示すとおり、エネルギーは増加し始める。同図のX軸は η 、Y軸は C_f である。

【0094】この方法で $C^{(n)}_f$ を最小にする最適な η の値を得ることができる。しかし、 λ の場合に比べていろいろな要素が計算に影響する結果、 $C^{(n)}_f$ は小

$$V(q_{f^{(m,s)}(i,j)}^{(m,s)})$$

(式36)

における輝度を持つ $f^{(m,s)}$ が提供される。つまりスーパーサンプリングが行われる。実験では、 f

$$V(q_{(i,j)+(0.5,0.5)}^{(m,s)})$$

(式37)

は、

【数38】

$$(V(q_{(i,j)}^{(m,s)}) + V(q_{(i,j)+(1,1)}^{(m,s)}))/2 \quad (\text{式38})$$

によって与えられた。

とが保障される。

【0091】なお、式34からわかるように、 r は画像の解像度に影響されること、すなわち r は 2^m に比例することに注意すべきである。このために [1. 4. 1] においてファクター 2^m を導入した。

【0092】 [1. 4. 3] η の動的決定

パラメータ η も同様の方法で自動決定できる。はじめに $\eta=0$ とし、最も細かい解像度における最終的な写像 $f^{(n)}$ 及びエネルギー $C^{(n)}_f$ を計算する。つづいて、 η をある値 $\Delta\eta$ だけ増加させ、再び最も細かい解像度における最終写像 $f^{(n)}$ 及びエネルギー $C^{(n)}_f$ を計算し直す。この過程を最適値が求まるまで続ける。 η は写像の剛性を示す。次式の重みだからである。

【0093】

【数35】

(式35)

さく揺らぎながら変化する。 λ の場合は、入力が微小量変化するたびに副写像を1回計算しなおすだけだが、 η の場合はすべての副写像が計算しなおされるためである。このため、得られた $C^{(n)}_f$ の値が最小であるかどうかを即座に判断することはできない。最小値の候補が見つければ、さらに細かい区間を設定することによって真の最小値を探す必要がある。

【0095】 [1. 5] スーパーサンプリング

画素間の対応関係を決定する際、自由度を増やすために、 $f^{(m,s)}$ の値域を $R \times R$ に拡張することができる (R は実数の集合)。この場合、終点画像の画素の輝度が補間され、非整数点、

【数36】

(m, s) は整数及び半整数値をとることが許され、

【数37】

【0096】 [1. 6] 各画像の画素の輝度の正規化

始点画像と終点画像がきわめて異なるオブジェクトを含んでいるとき、写像の計算に元の画素の輝度がそのままでは利用しにくい。輝度の差が大きいために輝度に関するエネルギー $C(m, s)_f$ が大きくなりすぎ、正しい評価がしづらいためである。

【0097】例えば、人の顔と猫の顔のマッチングをとる場合を考える。猫の顔は毛で覆われており、非常に明るい画素と非常に暗い画素が混じっている。この場合、ふたつの顔の間の副写像を計算するために、まず副画像を正規化する。すなわち、最も暗い画素の輝度を0、最も明るいそれを255に設定し、他の画素の輝度は線形補間によって求めておく。

【0098】[1. 7] インプリメンテーション
始点画像のスキャンに従って計算がリニアに進行する帰納的な方法を用いる。始めに、1番上の左端の画素

$(i, j) = (0, 0)$ について $f(m, s)$ の値を決定する。次に i を1ずつ増やしながら各 f

$(m, s)(i, j)$ の値を決定する。 i の値が画像の幅に到達したとき、 j の値を1増やし、 i を0に戻す。以降、始点画像のスキャンに伴い $f(m, s)(i, j)$ を決定していく。すべての点について画素の対応が決まれば、ひとつの写像 $f(m, s)$ が決まる。

【0099】ある $p(i, j)$ について対応点 $q_f(i, j)$ が決まれば、つぎに $p(i, j+1)$ の対応点 $q_f(i, j+1)$ が決められる。この際、 $q_f(i, j+1)$ の位置は全単射条件を満たすために、 $q_f(i, j)$ の位置によって制限される。したがって、先に対応点が決まる点ほどこのシステムでは優先度が高くなる。つねに $(0, 0)$ が最も優先される状態がつづくと、求められる最終の写像に余計な偏向が加わ

$$W = \vec{A} \times \vec{B}$$

ただしここで、

$$\vec{A} = \overrightarrow{q_{f(m,s)}^{(m,s)}(i,j-1) q_{f(m,s)}^{(m,s)}(i+1,j-1)}$$

【数41】

$$\vec{B} = \overrightarrow{q_{f(m,s)}^{(m,s)}(i,j-1) q_{f(m,s)}^{(m,s)}(k,l)}$$

である（ここでベクトルは三次元ベクトルとし、 z 軸は直交右手座標系において定義される）。もし W が負であれば、その候補については $D(m, s)(k, 1)$ に ψ を掛けることによってペナルティを与え、できるかぎり選択しないようにする。

【0104】図5(a)、図5(b)はこの条件を検査する理由を示している。図5(a)はペナルティのない候補、図5(b)はペナルティがある候補をそれぞれ表

る。本前提技術ではこの状態を回避するために、 $f(m, s)$ を以下の方法で決めていく。

【0100】まず $(s \bmod 4)$ が0の場合、 $(0, 0)$ を開始点とし i 及び j を徐々に増やしながら決めていく。 $(s \bmod 4)$ が1の場合、最上行の右端点を開始点とし、 i を減少、 j を増加させながら決めていく。 $(s \bmod 4)$ が2のとき、最下行の右端点を開始点とし、 i 及び j を減少させながら決めていく。 $(s \bmod 4)$ が3の場合、最下行の左端点を開始点とし、 i を増加、 j を減少させながら決めていく。解像度が最も細かい第 n レベルには副写像という概念、すなわちパラメータ s が存在しないため、仮に $s=0$ 及び $s=2$ であるとしてふたつの方向を連続的に計算した。

【0101】実際のインプリメンテーションでは、全単射条件を破る候補に対してペナルティを与えることにより、候補 $(k, 1)$ の中からできる限り全単射条件を満たす $f(m, s)(i, j)$ ($m=0, \dots, n$) の値を選んだ。第3の条件を破る候補のエネルギー $D(k, 1)$ には ϕ を掛け、一方、第1または第2の条件を破る候補には ψ を掛ける。今回は $\phi=2$ 、 $\psi=100000$ を用いた。

【0102】前述の全単射条件のチェックのために、実際の手続として $(k, 1) = f(m, s)(i, j)$ を決定する際に以下のテストを行った。すなわち $f(m, s)(i, j)$ の相続四辺形に含まれる各格子点 $(k, 1)$ に対し、次式の外積の z 成分が0以上になるかどうかを確かめる。

【0103】

【数39】

(式39)

【数40】

(式40)

(式41)

す。隣接画素 $(i, j+1)$ に対する写像 $f(m, s)(i, j+1)$ を決定する際、 W の z 成分が負であれば始点画像平面上において全単射条件を満たす画素は存在しない。なぜなら、 $q(m, s)(k, 1)$ は隣接する四辺形の境界線を越えるためである。

【0105】[1. 7. 1] 副写像の順序
インプリメンテーションでは、解像度レベルが偶数のときには $\sigma(0)=0$ 、 $\sigma(1)=1$ 、 $\sigma(2)=2$ 、 σ

(3) = 3、 $\sigma(4) = 0$ を用い、奇数のときは $\sigma(0) = 3$ 、 $\sigma(1) = 2$ 、 $\sigma(2) = 1$ 、 $\sigma(3) = 0$ 、 $\sigma(4) = 3$ を用いた。このことで、副写像を適度にシャッフルした。なお、本来副写像は4種類であり、 s は0～3のいずれかである。しかし、実際には $s = 4$ に相当する処理を行った。その理由は後述する。

【0106】[1. 8] 補間計算

始点画像と終点画像の間の写像が決定された後、対応しあう画素の輝度が補間される。実験では、トライリニア補間を用いた。始点画像平面における正方形 p

$$(i, j) \quad p(i+1, j) \quad p(i, j+1) \quad p$$

$$\begin{aligned} (x, y) = & (1-dx)(1-dy)(1-t)(i, j) + (1-dx)(1-dy)tf(i, j) \\ & + dx(1-dy)(1-t)(i+1, j) + dx(1-dy)tf(i+1, j) \\ & + (1-dx)dy(1-t)(i, j+1) + (1-dx)dytf(i, j+1) \\ & + dxdy(1-t)(i+1, j+1) + dxdytf(i+1, j+1) \end{aligned}$$

つづいて $r(x, y, t)$ における画素の輝度が次の式を用いて決定される。

$(i+1, j+1)$ が終点画像平面上の四辺形 q
 $r(i, j) \quad q_r(i+1, j) \quad q_r(i, j+1) \quad q_r(i+1, j+1)$ に射影されると仮定する。簡単のため、画像間の距離を1とする。始点画像平面からの距離が t ($0 \leq t \leq 1$) である中間画像の画素 $r(x, y, t)$ ($0 \leq x \leq N-1, 0 \leq y \leq M-1$) は以下の要領で求められる。まず画素 $r(x, y, t)$ の位置 (ただし $x, y, t \in R$) を次式で求める。

【0107】

【数42】

(式42)

【0108】

【数43】

$$\begin{aligned} V(r(x, y, t)) = & (1-dx)(1-dy)(1-t)V(p_{(i,j)}) + (1-dx)(1-dy)tV(q_{f(i,j)}) \\ & + dx(1-dy)(1-t)V(p_{(i+1,j)}) + dx(1-dy)tV(q_{f(i+1,j)}) \\ & + (1-dx)dy(1-t)V(p_{(i,j+1)}) + (1-dx)dytV(q_{f(i,j+1)}) \\ & + dxdy(1-t)V(p_{(i+1,j+1)}) + dxdytf(q_{f(i+1,j+1)}) \end{aligned}$$

(式43)

ここで dx 及び dy はパラメータであり、0から1まで変化する。

【0109】[1. 9] 拘束条件を課したときの写像
 いままでは拘束条件がいっさい存在しない場合の写像の決定を述べた。しかし、始点画像と終点画像の特定の画素間に予め対応関係が規定されているとき、これを拘束条件としたうえで写像を決定することができる。

【0110】基本的な考えは、まず始点画像の特定の画素を終点画像の特定の画素に移す大まかな写像によって始点画像を大まかに変形し、しかる後、写像 f を正確に計算する。

$$p(i_0, j_0), p(i_1, j_1), \dots,$$

$$p(i_{n_s-1}, j_{n_s-1})$$

(式44)

を特定するとき、以下の値を決める。

【数45】

【0111】まず始めに、始点画像の特定の画素を終点画像の特定の画素に射影し、始点画像の他の画素を適当な位置に射影する大まかな写像を決める。すなわち、特定の画素に近い画素は、その特定の画素が射影される場所の近くに射影されるような写像である。ここで第 m レベルの大まかな写像を $F(m)$ と記述する。

【0112】大まかな写像 F は以下の要領で決める。まず、いくつかの画素について写像を特定する。始点画像について n_s 個の画素、

【数44】

$$F^{(n)}(i_0, j_0) = (k_0, l_0),$$

$$F^{(n)}(i_1, j_0) = (k_1, l_1), \dots,$$

$$F^{(n)}(i_{n_s-1}, j_{n_s-1}) = (k_{n_s-1}, l_{n_s-1})$$

(式45)

始点画像の他の画素の変位量は、 $p(i_h, j_h)$ ($h = 0, \dots, n_s - 1$) の変位に重み付けをして求められる平均である。すなわち画素 $p(i, j)$ は、終点画像

の以下の画素に射影される。

【0113】

【数46】

$$F^{(m)}(i, j) = \frac{(i, j) + \sum_{h=0}^{n_s-1} (k_h - i_h, l_h - j_h) \text{weight}_h(i, j)}{2^{n-m}}$$

(式46)

ただしここで、

【数47】

$$\text{weight}_h(i, j) = \frac{1/\|(i_h - i, j_h - j)\|^2}{\text{total weight}(i, j)}$$

(式47)

【数48】

$$\text{total weight}(i, j) = \sum_{h=0}^{n_s-1} 1/\|(i_h - i, j_h - j)\|^2$$

(式48)

とする。

ギード (m, s) (i, j) を変更する。正確には、 D

【0114】 つづいて、 $F^{(m)}$ に近い候補写像 f がより少ないエネルギーを持つように、その写像 f のエネルギー

(m, s) (i, j) は、

【数49】

$$D_{(i,j)}^{(m,s)} = E_{0(i,j)}^{(m,s)} + \eta E_{1(i,j)}^{(m,s)} + \kappa E_{2(i,j)}^{(m,s)}$$

(式49)

である。ただし、

【数50】

$$E_{2(i,j)}^{(m,s)} = \begin{cases} 0, & \text{if } \|F^{(m)}(i, j) - f^{(m,s)}(i, j)\|^2 \leq \lfloor \frac{\rho^2}{2^{2(n-m)}} \rfloor \\ \|F^{(m)}(i, j) - f^{(m,s)}(i, j)\|^2, & \text{otherwise} \end{cases}$$

(式50)

であり、 $\kappa, \rho \geq 0$ とする。最後に、前述の写像の自動計算プロセスにより、 f を完全に決定する。

【0115】 ここで、 $f^{(m, s)}(i, j)$ が $F^{(m)}(i, j)$ に十分近いとき、つまりそれらの距離が、

【数51】

$$\left\lfloor \frac{\rho^2}{2^{2(n-m)}} \right\rfloor$$

(式51)

以内であるとき、 $E_{2(i,j)}^{(m, s)}(i, j)$ が0になることに注意すべきである。そのように定義した理由は、各 $f^{(m, s)}(i, j)$ が $F^{(m)}(i, j)$ に十分近い限り、終点画像において適切な位置に落ち着くよう、その値を自動的に決めたいためである。この理由により、正

確な対応関係を詳細に特定する必要がなく、始点画像は終点画像にマッチするように自動的にマッピングされる。

【0116】 [2] 具体的な処理手順

[1] の各要素技術による処理の流れを説明する。図6は前提技術の全体手順を示すフローチャートである。同図のごとく、まず多重解像度特異点フィルタを用いた処理を行い (S1)、つづいて始点画像と終点画像のマッチングをとる (S2)。ただし、S2は必須ではなく、S1で得られた画像の特徴をもとに画像認識などの処理を行ってもよい。

【0117】 図7は図6のS1の詳細を示すフローチャートである。ここではS2で始点画像と終点画像のマッ

チングをとることを前提としている。そのため、まず特異点フィルタによって始点画像の階層化を行い（S10）、一連の始点階層画像を得る。つづいて同様の方法で終点画像の階層化を行い（S11）、一連の終点階層画像を得る。ただし、S10とS11の順序は任意であるし、始点階層画像と終点階層画像を並行して生成していくこともできる。

【0118】図8は図7のS10の詳細を示すフローチャートである。もとの始点画像のサイズは $2n \times 2n$ とする。始点階層画像は解像度が細かいほうから順に作られるため、処理の対象となる解像度レベルを示すパラメータ m を n にセットする（S100）。つづいて第 m レベルの画像 $p(m, 0)$ 、 $p(m, 1)$ 、 $p(m, 2)$ 、 $p(m, 3)$ から特異点フィルタを用いて特異点を検出し（S101）、それぞれ第 $m-1$ レベルの画像 $p(m-1, 0)$ 、 $p(m-1, 1)$ 、 $p(m-1, 2)$ 、 $p(m-1, 3)$ を生成する（S102）。ここでは $m=n$ であるため、 $p(m, 0) = p(m, 1) = p(m, 2) = p(m, 3) = p(n)$ であり、ひとつの始点画像から4種類の副画像が生成される。

【0119】図9は第 m レベルの画像の一部と、第 $m-1$ レベルの画像の一部の対応関係を示している。同図の数値は各画素の輝度を示す。同図の $p(m, s)$ は $p(m, 0) \sim p(m, 3)$ の4つの画像を象徴するもので、 $p(m-1, 0)$ を生成する場合には、 $p(m, s)$ は $p(m, 0)$ であると考え、[1, 2]で示した規則により、 $p(m-1, 0)$ は例えば同図で輝度を記入したブロックについて、そこに含まれる4画

$$\sum_s \sum_j (\lambda C(m, s)(i, j) + \eta E_0(m, s)(i, j) + E_1(m, s)(i, j)) \quad (\text{式52})$$

となる。ただし、総和は i, j についてそれぞれ0, 1, ..., $2m-1$ で計算する。以上でマッチング評価の準備が整う。

【0123】図12は図6のS2の詳細を示すフローチャートである。[1]で述べたごとく、始点階層画像と終点階層画像のマッチングは互いに同じ解像度レベルの画像どうしでとられる。画像間の大域的なマッチングを良好にとるために、解像度が粗いレベルから順にマッチングを計算する。特異点フィルタを用いて始点階層画像および終点階層画像を生成しているため、特異点の位置や輝度は解像度の粗いレベルでも明確に保存されており、大域的なマッチングの結果は従来に比べて非常に優れたものになる。

【0124】図12のごとく、まず係数パラメータ η を0、レベルパラメータ m を0に設定する（S20）。つづいて、始点階層画像中の第 m レベルの4つの副画像と終点階層画像中の第 m レベルの4つの副画像のそれぞれの間でマッチングを計算し、それぞれ全単射条件を満たし、かつエネルギーを最小にするような4種類の副画像

素のうち「3」、 $p(m-1, 1)$ は「8」、 $p(m-1, 2)$ は「6」、 $p(m-1, 3)$ を「10」をそれぞれ取得し、このブロックをそれぞれ取得したひとつの画素で置き換える。したがって、第 $m-1$ レベルの副画像のサイズは $2m-1 \times 2m-1$ になる。

【0120】つづいて m をデクリメントし（図8のS103）、 m が負になっていないことを確認し（S104）、S101に戻ってつぎに解像度の粗い副画像を生成していく。この繰り返し処理の結果、 $m=0$ 、すなわち第0レベルの副画像が生成された時点でS10が終了する。第0レベルの副画像のサイズは 1×1 である。

【0121】図10はS10によって生成された始点階層画像を $n=3$ の場合について例示している。最初の始点画像のみが4つの系列に共通であり、以降特異点の種類に応じてそれぞれ独立に副画像が生成されていく。なお、図8の処理は図7のS11にも共通であり、同様の手順を経て終点階層画像も生成される。以上で図6のS1による処理が完了する。

【0122】前提技術では、図6のS2に進むためにマッチング評価の準備をする。図11はその手順を示している。同図のごとく、まず複数の評価式が設定される（S30）。[1, 3, 2, 1]で導入した画素に関するエネルギー $C(m, s)_f$ と[1, 3, 2, 2]で導入した写像の滑らかさに関するエネルギー $D(m, s)_f$ がそれぞれである。つぎに、これらの評価式を統合して総合評価式を立てる（S31）。[1, 3, 2, 3]で導入した総エネルギー $\lambda C(m, s)_f + D(m, s)_f$ がそれぞれであり、[1, 3, 2, 2]で導入した η を用い

れば、 $f(m, s)$ （ $s=0, 1, 2, 3$ ）を求める（S21）。全単射条件は[1, 3, 3]で述べた相続四辺形を用いて検査される。この際、式17、18が示すように、第 m レベルにおける副写像は第 $m-1$ レベルのそれらに拘束されるため、より解像度の粗いレベルにおけるマッチングが順次利用されていく。これは異なるレベル間の垂直的参照である。なお、いま $m=0$ であってそれより粗いレベルはないが、この例外的な処理は図13で後述する。

【0125】一方、同一レベル内における水平的参照も行われる。[1, 3, 3]の式20のごとく、 $f(m, 3)$ は $f(m, 2)$ に、 $f(m, 2)$ は $f(m, 1)$ に、 $f(m, 1)$ は $f(m, 0)$ に、それぞれ類似するように決める。その理由は、特異点の種類が違って、それらがもともとと同じ始点画像と終点画像に含まれている以上、副写像がまったく異なるという状況は不自然だからである。式20からわかるように、副写像どうしが近いほどエネルギーは小さくなり、マッチングが良好とみなされる。

【0126】なお、最初に決めるべき $f(m, 0)$ については同一のレベルで参照できる副写像がないため、式19に示すごとくひとつ粗いレベルを参照する。ただし、実験では $f(m, 3)$ まで求めた後、これを拘束条件として $f(m, 0)$ を一回更新するという手続をとった。これは式20に $s=4$ を代入し、 $f(m, 4)$ を新たな $f(m, 0)$ とすることに等しい。 $f(m, 0)$ と $f(m, 3)$ の関連度が低くなり過ぎる傾向を回避するためであり、この措置によって実験結果がより良好になった。この措置に加え、実験では [1, 7, 1] に示す副写像のシャッフルも行った。これも本来特異点の種類ごとに決まる副写像どうしの関連度を密接に保つ趣旨である。また、処理の開始点に依存する偏向を回避するために、 s の値にしたがって開始点の位置を変える点は [1, 7] で述べたとおりである。

【0127】図13は第0レベルにおいて副写像を決定する様子を示す図である。第0レベルでは各副画像がただひとつの画素で構成されるため、4つの副写像 $f(0, s)$ はすべて自動的に単位写像に決まる。図14は第1レベルにおいて副写像を決定する様子を示す図である。第1レベルでは副画像がそれぞれ4画素で構成される。同図ではこれら4画素が実線で示されている。いま、 $p(1, s)$ の点 x の対応点を $q(1, s)$ の中で探すとき、以下の手順を踏む。

- 【0128】1. 第1レベルの解像度で点 x の左上点 a 、右上点 b 、左下点 c 、右下点 d を求める。
2. 点 $a \sim d$ がひとつ粗いレベル、つまり第0レベルにおいて属する画素を探す。図14の場合、点 $a \sim d$ はそれぞれ画素 $A \sim D$ に属する。ただし、画素 $A \sim C$ は本来存在しない仮想的な画素である。
3. 第0レベルですでに求まっている画素 $A \sim D$ の対応点 $A' \sim D'$ を $q(1, s)$ の中にプロットする。画素 $A' \sim C'$ は仮想的な画素であり、それぞれ画素 $A \sim C$ と同じ位置にあるものとする。
4. 画素 A の中の点 a の対応点 a' が画素 A' の中にあるとみなし、点 a' をプロットする。このとき、点 a が画素 A の中で占める位置（この場合、右下）と、点 a' が画素 A' の中で占める位置が同じであると仮定する。
5. 4と同様の方法で対応点 $b' \sim d'$ をプロットし、点 $a' \sim d'$ で相続四辺形を作る。
6. 相続四辺形の中でエネルギーが最小になるよう、点 x の対応点 x' を探す。対応点 x' の候補として、例えば画素の中心が相続四辺形に含まれるものに限定してもよい。図14の場合、4つの画素がすべて候補になる。

【0129】以上がある点 x の対応点の決定手順である。同様の処理を他のすべての点について行い、副写像を決める。第2レベル以上のレベルでは、次第に相続四辺形の形が崩れていくと考えられるため、図3に示すように画素 $A' \sim D'$ の間隔が空いていく状況が発生する。

【0130】こうして、ある第 m レベルの4つの副写像が決まれば、 m をインクリメントし（図12のS22）、 m が n を超えていないことを確かめて（S23）、S21に戻る。以下、S21に戻るたびに次第に細かい解像度のレベルの副写像を求め、最後にS21に戻ったときに第 n レベルの写像 $f(n)$ を決める。この写像は $\eta=0$ に関して定まったものであるから、 $f(n)$ ($\eta=0$) と書く。

【0131】つぎに異なる η に関する写像も求めるべく、 η を $\Delta\eta$ だけシフトし、 m をゼロクリアする（S24）。新たな η が所定の探索打ち切り値 η_{max} を超えていないことを確認し（S25）、S21に戻り、今回の η に関して写像 $f(n)$ ($\eta=\Delta\eta$) を求める。この処理を繰り返し、S21で $f(n)$ ($\eta=i\Delta\eta$) ($i=0, 1, \dots$) を求めていく。 η が η_{max} を超えたときS26に進み、後述の方法で最適な $\eta=\eta_{opt}$ を決定し、 $f(n)$ ($\eta=\eta_{opt}$) を最終的に写像 $f(n)$ とする。

【0132】図15は図12のS21の詳細を示すフローチャートである。このフローチャートにより、ある定まった η について、第 m レベルにおける副写像が決まる。副写像を決める際、前提技術では副写像ごとに最適な λ を独立して決める。

【0133】同図のごとく、まず s と λ をゼロクリアする（S210）。つぎに、そのときの λ について（および暗に η について）エネルギーを最小にする副写像 $f(m, s)$ を求め（S211）、これを $f(m, s)$ ($\lambda=0$) と書く。異なる λ に関する写像も求めるべく、 λ を $\Delta\lambda$ だけシフトし、新たな λ が所定の探索打ち切り値 λ_{max} を超えていないことを確認し（S213）、S211に戻り、以降の繰り返し処理で $f(m, s)$ ($\lambda=i\Delta\lambda$) ($i=0, 1, \dots$) を求める。 λ が λ_{max} を超えたときS214に進み、最適な $\lambda=\lambda_{opt}$ を決定し、 $f(m, s)$ ($\lambda=\lambda_{opt}$) を最終的に写像 $f(m, s)$ とする（S214）。

【0134】つぎに、同一レベルにおける他の副写像を求めるべく、 λ をゼロクリアし、 s をインクリメントする（S215）。 s が4を超えていないことを確認し（S216）、S211に戻る。 $s=4$ になれば上述のごとく $f(m, 3)$ を利用して $f(m, 0)$ を更新し、そのレベルにおける副写像の決定を終了する。

【0135】図16は、ある m と s について λ を変えながら求められた $f(m, s)$ ($\lambda=i\Delta\lambda$) ($i=0, 1, \dots$) に対応するエネルギー $C(m, s)_f$ の挙動を示す図である。[1, 4] で述べたとおり、 λ が増加すると通常 $C(m, s)_f$ は減少する。しかし、 λ が最適値を超えると $C(m, s)_f$ は増加に転じる。そこで本前提技術では $C(m, s)_f$ が極小値をとるときの λ を λ_{opt} と決める。同図のように $\lambda > \lambda_{opt}$ の範囲で再度 $C(m, s)_f$ が小さくなっていても、その時点

ではすでに写像がくずれていて意味をなさないため、最初の極小点に注目すればよい。 λ_{opt} は副写像ごとに独立して決めていき、最後に $f(n)$ についてもひとつ定まる。

【0136】一方、図17は、 η を変えながら求められた $f(n)$ ($\eta = i \Delta \eta$) ($i = 0, 1, \dots$)に対応するエネルギー $C(n)_f$ の挙動を示す図である。ここでも η が増加すると通常 $C(n)_f$ は減少するが、 η が最適値を超えると $C(n)_f$ は増加に転じる。そこで $C(n)_f$ が極小値をとるときの η を η_{opt} と決める。図17は図4の横軸のゼロ付近を拡大した図と考えてよい。 η_{opt} が決まれば $f(n)$ を最終決定することができる。

【0137】以上、本前提技術によれば種々のメリットが得られる。まずエッジを検出する必要がないため、エッジ検出タイプの従来技術の課題を解消できる。また、画像に含まれるオブジェクトに対する先験的な知識も不要であり、対応点の自動検出が実現する。特異点フィルタによれば、解像度の粗いレベルでも特異点の輝度や位置を維持することができ、オブジェクト認識、特徴抽出、画像マッチングに極めて有利である。その結果、人手作業を大幅に軽減する画像処理システムの構築が可能となる。

【0138】なお、本前提技術について次のような変形技術も考えられる。

(1) 前提技術では始点階層画像と終点階層画像の間でマッチングをとる際にパラメータの自動決定を行ったが、この方法は階層画像間ではなく、通常の2枚の画像間のマッチングをとる場合全般に利用できる。

【0139】たとえば2枚の画像間で、画素の輝度の差に関するエネルギー E_0 と画素の位置的なずれに関するエネルギー E_1 のふたつを評価式とし、これらの線形和 $E_{tot} = \alpha E_0 + E_1$ を総合評価式とする。この総合評価式の極値付近に注目して α を自動決定する。つまり、いろいろな α について E_{tot} が最小になるような写像を求める。それらの写像のうち、 α に関して E_1 が極小値をとるときの α を最適パラメータと決める。そのパラメータに対応する写像を最終的に両画像間の最適マッチングとみなす。

【0140】これ以外にも評価式の設定にはいろいろな方法があり、例えば $1/E_1$ と $1/E_2$ のように、評価結果が良好なほど大きな値をとるものを採用してもよい。総合評価式も必ずしも線形和である必要はなく、 n 乗和($n = 2, 1/2, -1, -2$ など)、多項式、任意の関数などを適宜選択すればよい。

【0141】パラメータも、 α のみ、前提技術のごとく η と λ のふたつの場合、それ以上の場合など、いずれでもよい。パラメータが3以上の場合はひとつずつ変化させて決めていく。

【0142】(2) 本前提技術では、総合評価式の値が

最小になるよう写像を決めた後、総合評価式を構成するひとつの評価式である $C(m, s)_f$ が極小になる点を検出してパラメータを決定した。しかし、こうした二段回処理の代わりに、状況によっては単に総合評価式の最小値が最小になるようにパラメータを決めても効果的である。その場合、例えば $\alpha E_0 + \beta E_1$ を総合評価式とし、 $\alpha + \beta = 1$ なる拘束条件を設けて各評価式を平等に扱うなどの措置を講じてよい。パラメータの自動決定の本質は、エネルギーが最小になるようにパラメータを決めていく点にあるからである。

【0143】(3) 前提技術では各解像度レベルで4種類の特異点に関する4種類の副画像を生成した。しかし、当然4種類のうち1、2、3種類を選択的に用いてもよい。例えば、画像中に明るい点がひとつだけ存在する状態であれば、極大点に関する $f(m, 3)$ だけで階層画像を生成しても相応の効果を得られるはずである。その場合、同一レベルで異なる副写像は不要になるため、 s に関する計算量が減る効果がある。

【0144】(4) 本前提技術では特異点フィルタによってレベルがひとつ進むと画素が $1/4$ になった。例えば 3×3 で1ブロックとし、その中で特異点を探す構成も可能であり、その場合、レベルがひとつ進むと画素は $1/9$ になる。

【0145】(5) 始点画像と終点画像がカラーの場合、それらをまず白黒画像に変換し、写像を計算する。その結果求められた写像を用いて始点のカラー画像を変換する。それ以外の方法として、RGBの各成分について副写像を計算してもよい。

【0146】[映像効果または画像補間に関する実施の形態] 以上の前提技術を利用した実施の形態を説明する。実施の形態1はより原理に近いもの、実施の形態2以降はよりアプリケーションに近いものを挙げるが、それらに明確な境界線を引く必要はなく、いずれかの実施の形態の任意の機能と他の実施の形態の任意の機能が組み合わせられてもよい。

【0147】実施の形態1

図18は、キーフレームである第1画像 I_1 、第2画像 I_2 を示し、それらの間で、ある画素 $p_1(x_1, y_1)$ と $p_2(x_2, y_2)$ が対応している。これらの対応は前提技術で求められている。

【0148】図19は、第1画像 I_1 上にメッシュを切り、そのメッシュを構成するポリゴンの第2画像 I_2 上の対応位置を示す。いま第1画像 I_1 上で注目するポリゴン R_1 は4個の格子点A、B、C、Dによって決まっている。このポリゴン R_1 を「起点ポリゴン」と名付ける。これら4個の格子点A、B、C、Dは、図18に示したごとく、それぞれ第2画像 I_2 上に対応する点 A' 、 B' 、 C' 、 D' を有し、これらの点で形成されるポリゴン R_2 を「終点ポリゴン」と名付ける。起点ポリゴンは一般に長方形であるが、終点ポリゴンは一般に

四辺形である。いずれにせよ、この実施の形態では、画素単位で第1画像I1と第2画像I2の対応関係を記述せず、起点ポリゴンの格子点について、その対応画素を記述する。その記述は対応点ファイルに書き込まれる。格子点に注目することにより、対応点ファイルの容量を非常に小さくすることができる。

【0149】対応点ファイルは第1画像I1と第2画像I2の中間画像を生成するために利用される。この点は前提技術で述べたごとく、対応点どうしの位置を補間することにより、任意の時間的位置の中間画像を生成できる。したがって、第1画像I1、第2画像I2、および対応点ファイルを保存しておくことで、それに2つの画像間のモーフィングや滑らかな動画を生成することができる。このため、動画の圧縮効果が得られる。

【0150】図20は対応点ファイルから格子点以外の点に関する対応関係を計算する方法を示す。対応点ファイルには格子点に関する情報しかないので、ポリゴン内部の点について対応情報を別途計算する必要がある。図20では、図19の起点ポリゴンR1の下半分にあたる三角形ABCと、同様に終点ポリゴンR2の下半分にあたる三角形A'B'C'との対応を示す。いま、起点ポリゴン側の三角形ABC内部の点Qが、線分ACを $t:(1-t)$ に内分し、その内分点と点Bの間を $s:(1-s)$ に内分する位置にあったとする。この点Qの終点ポリゴン側の三角形A'B'C'における対応点Q'は、線分A'C'を $t:(1-t)$ に内分し、その内分点と点B'の間を $s:(1-s)$ に内分する位置にあると考えればよい。すなわち、起点ポリゴンを三角形に分割し、その三角形に関するベクトルの内分の形式にて終点ポリゴン内部の点を求めればよい。ベクトル斜体で示せば、

$$\vec{BQ} = (1-s) \{ (1-t) \vec{BA} + t \vec{BC} \}$$

であり、したがって、

$$\vec{B'Q'} = (1-s) \{ (1-t) \vec{B'A'} + t \vec{B'C'} \}$$

当然ながら、同様の処理を起点ポリゴンR1の上半分にあたる三角形ACDと、同様に終点ポリゴンR2の上半分にあたる三角形A'C'D'との間についてもなすとする。

【0151】図21は、以上の処理手順を示す。まず図19のごとく、第1画像I1上に取られた格子点に関するマッチング結果を取得する(S10)。このとき、前提技術の画素単位のマッチングを実行し、その結果の中から格子点に当たる部分を抽出すればよい。ただし、前提技術を用いずに他のマッチング技術、例えばオブティカルフローやブロックマッチングをもとに格子点に関するマッチング結果を特定してもよい。

【0152】つづいて、図19の右側のごとく、第2画像I2上に終点ポリゴンを定義する(S12)。以上で対応点ファイルが生成できるので、このファイルに第1

画像I1、第2画像I2を識別するための情報を盛り込んで出力する(S14)。ふたつの画像および対応点ファイルは、任意の記録装置または媒体に保持され、または直接ネットワークや放送波に乗せて伝送されればよい。

【0153】図22は、対応点ファイルを用いて中間画像を生成する手順を示す。まず、第1画像I1、第2画像I2が読み込まれ(S20)、対応点ファイルが読み込まれる(S22)。つづいて、起点ポリゴン内の点と終点ポリゴン内の点の対応関係が図20の方法によって算出される(S24)。この時点で、画像内のすべての画素に関する対応関係を取得することができる。したがって、前提技術で述べたごとく、互いに対応し合う点の座標と色を $u:(1-u)$ で内分すれば、第1画像I1と第2画像I2の間を時間的に $(1-u):u$ で内分する位置における中間画像を生成することができる(S26)。なお、内挿補間だけでなく、外挿補間をしてもよい。

【0154】図23は以上の処理を行う映像効果装置10の構成を示す。この装置10は、外部の記憶装置や撮影カメラ等から第1画像I1、第2画像I2を取得する画像入力部12と、それらの画像に前提技術その他の技術によってマッチング計算を施すマッチングプロセッサ14と、それによって生成された対応点ファイルFを格納する対応点ファイル保持部16と、第1画像I1および第2画像I2から中間画像を生成する中間画像生成部18と、第1画像I1、第2画像I2および中間画像をタイミング調整しながらもとの動画に近い映像として表示する表示部20を含む。また、通信部22は、外部からの要望にしたがい、第1画像I1、第2画像I2および対応点ファイルFをネットワークその他の伝送インフラへ送出する。なお、マッチングプロセッサ14には、メッシュの大きさや格子点の位置などを示すメッシュ情報が入力されている。

【0155】以上の構成により、画像入力部12に入力された第1画像I1と第2画像I2はマッチングプロセッサ14へ送られる。マッチングプロセッサ14は、それらの画像間で画素単位のマッチング計算を行う。マッチングプロセッサ14はメッシュ情報をもとに対応点ファイルFを生成し、対応点ファイル保持部16へ出力する。

【0156】中間画像生成部18は、ユーザの求めに応じ、またはその他の要因で対応点ファイルFを読み出し、中間画像を生成する。この中間画像は表示部20へ送られ、そこで画像出力の時間調整が行われ、動画またはモーフィング画像が表示される。この動作からもわかるとおり、中間画像生成部18と表示部20は、本装置10とは別構成のリモート端末側にあってもよく、その場合、その端末が第1画像I1、第2画像I2、対応点ファイルFという比較的軽いデータを受け、自主的に動

画を再生できる。

【0157】通信部22はそのための構成であり、ここではすでにリモート端末がネットワークの先で待っていることを想定している。すなわち、通信部22は第1画像I1、第2画像I2、対応点ファイルFをネットワークや放送波に乗せて送り出し、リモート端末側で動画の表示が行われる。もちろんリモート端末は、表示ではなく、ストレージを目的としてもよい。なお本装置は、ネットワーク等の外部から通信部22を介して第1画像I1、第2画像I2およびそれらの対応点ファイルを入力し、これらを中間画像生成部18に渡して補間計算をし、中間画像を生成する用途もある。図23にはそのためのデータの経路Pが記述されている。

【0158】実施の形態1の処理内容について実験を行ったところ、第1、第2画像として、例えば 256×256 程度の大きさを採用し、格子点を縦横10～数十画素おき程度に設定することにより、きわめて良好なモーフニングまたは動画圧縮効果が得られた。対応点ファイルの大きさは数キロから10キロバイト程度であり、画質の高さとデータ量の小ささが実現できることが確認できた。

【0159】実施の形態2

実施の形態1で実現された映像効果装置は、電子商取引における商品プレゼンテーションに利用できる。実施の形態2のこの装置（以下「商品提示装置」という）は、いくつかの角度から撮影された商品の写真のうち近接しあう2枚の画像を第1画像I1、第2画像I2とし、それらの間を滑らかに補間することで商品をユーザに示す。たとえば、 10° おきに商品のまわりから撮影した写真を順次補間することで 360° 滑らかな回転画像が得られる。従来、商品の滑らかな回転画像はせいぜい 1° 、 2° ごとに撮影が必要とされたが、実施の形態1の技術を利用することで、この角度を $10 \sim 20^\circ$ にできることが判明している。

【0160】商品提示装置は、実施の形態1と異なり、予め生成された対応点ファイルFを利用する。商品の写真は提示前に準備できるためである。図23の構成との相違点のみ示せば、画像入力部12は第1画像I1、第2画像I2のほかに、対応点ファイルFを取得する。これらは中間画像生成部18へ入力され、中間画像が生成される。中間画像は表示部20で表示されたり、通信部22を介してネットワークその他の伝送インフラへ送出されてもよい。この実施の形態ではマッチングプロセッサ14と対応点ファイル保持部16は不要である。なお、そうした構成は他の実施の形態にも有効である。

【0161】商品提示装置は、実施の形態1に加え、電子商取引における商品プレゼンテーション機能を提供する。この機能が前述の商品回転機能である場合、ユーザからGUIを通じて入力された指示にしたがって商品を回転させてもよい。そうした指示の方法は、たとえばC

Gの分野でいろいろと知られており、ここではそうした技術を利用する。

【0162】機能の別の例として、商品画像を二方向に補間して得られた中間画像を表示してもよい。たとえば、商品を横方向にとびとびに撮影することにくわえ、縦方向にもとびとびに撮影することにより、横方向だけでなく、縦方向にも補間することができる。この機能は例えば下の工程を含む。

(1) 第1画像、第2画像からなる画像対Aと、それら2枚の画像間の対応点情報Aを取得する工程

(2) 第3画像、第4画像からなる画像対Bと、それら2枚の画像間の対応点情報Bを取得する工程

(3) 画像対Aの2枚の画像間に時間的または空間的に定まる軸Aと画像対Bの2枚の画像間に時間的または空間的に定まる軸Bの位置関係、対応点情報A、および対応点情報Bを利用して、中間画像を補間によって生成する工程。

【0163】(3)において、軸Aと軸Bを利用して2軸補間(bilinear interpolation)を行ってもよい。一例として、2つの視点 $p_1(0, 0)$ 、 $p_2(0, 100)$ から得た画像対Aとし、別の2つの視点 $p_3(100, 0)$ 、 $p_4(100, 100)$ から得た画像対Bとする。点 p_1 と p_2 を結ぶ直線が軸A、点 p_3 と p_4 を結ぶ直線が軸Bに当たる。

【0164】中間画像として、視点 $p' = (50, 50)$ から見た画像を得る場合、まず画像対Aの間の対応点情報Aをもとに視点 $(0, 50)$ からのフレームを生成する。つぎに画像対Bの間の対応点情報Bをもとに視点 $(100, 50)$ からの別のフレームを生成する。つづいて、これらふたつのフレームに補間を施し、すなわちこの場合1:1で内分し、所望の中間フレームが生成される。ここで、縦横両方向の補間をするためには、一般には、軸Aと軸Bが同一直線上にこないよう画像対Aと画像対Bを定めればよい。

【0165】なお、この例では軸A、軸Bは、それぞれ2枚の画像間に空間的に定まるが、時間的に定まる別の例もある。たとえば、ひとつの視点Pから時刻 $t = t_0$ と t_1 に得たふたつの画像を画像対Aとし、別の視点Qから時刻 $t = t_0$ と t_1 に得たふたつの画像を画像対Bとする。この場合、画像対Aにおいて (P, t_0) で定義される点と (P, t_1) で定義される点を結ぶ直線が軸Aとなり、同様に画像対Bにおいて (Q, t_0) で定義される点と (Q, t_1) で定義される点を結ぶ直線が軸Bとなる。したがって、中間画像を例えば点 $((P+Q)/2, (t_0+t_1)/2)$ からの画像とすれば、それぞれふたつの軸の中間的な画像を生成した後、これらを補間すればよい。

【0166】実施の形態3

中間画像の生成に当たり、映像効果に関する機能を提供してもよい。実施の形態1では一般論としての映像効果

に触れたが、ここでいう映像効果は、より具体的またはプロフェッショナルに近い用途と考えてもよい。以下機能の例である。

【0167】3以上の画像の中から隣接するふたつの画像を第1画像および第2画像として選定し、それらの中間画像を生成および表示する。つづいて、別の隣接する画像を第1画像と第2画像とし、以降順に中間画像を生成する。例えば、視点が少しずつ異なる画像S、T、Uがあるとき、まず画像SとTの中間画像を作り、つぎに画像TとUの中間画像を作ることにより、一連の画像を連続的に表示することができる。すなわち、多数の離散的な視点画像間をつぎつぎ補間して中間画像を生成すれば、あたかもパノラマ的な画像が得られる。

【0168】これと同様の方法であるが、時間の要素を取り入れることもできる。例えば、複数の離散的な画像を、視点を時間とともに動かした画像と捉え、かつそれらの中間画像を順を追って表示すれば、いわゆるウォークスルー画像が得られる。

【0169】別の機能は、処理の対象となる画像を二方向に補間して得られた中間画像を表示するものである。この方法は実施の形態2でも述べたとおり、例えば二軸補間で実現する。これにより、例えば二方向に連続変化するパノラマ画像も得られる。

【0170】さらに別の機能は、ある対象物を少しずつ廻りながら撮影し、それらの連続画像を生成する機能である。実施の形態2では商品自体を回転させたが、映像効果としては、むしろ対象物をぐるりと撮影した画像を表示する点におもしろさがある。一部の映画では、空中にとどまった人を一周する映像がSFX技術で生成されているが、実施の形態1によれば、より少ない写真画像で容易にそうした画像を生成することができる。

【0171】これら以外にも既存の映像効果を実施の形態に係る技術で効率化し、または高画質化することができる。本実施の形態に係る装置は、いずれの映像効果を行うかをユーザが指定するGUIを搭載し、指定内容に相応しい前処理または後処理の機能を実装しておいてもよい。

【0172】たとえば、ウォークスルー画像を生成するとき、予め視点の移動経路を指定してもよく、それにしたがって順に中間画像を表示してもよい。二方向に補間する場合は、実施の形態2同様、GUI、例えばマウスの移動方向にしたがって移動する視点を仮定し、補間画像を生成してもよい。マウスが右下から左上へ移動すれば、視点も右下から左上へ移動したと想定し、二軸補間をすればよい。

【0173】実施の形態4

実施の形態1でも触れたごとく、本発明は動画の圧縮にも向く。少ない静止画像をもとに動画が再生できるため、インターネット等で動画を配信する際に有利である。このため、中間画像の生成に当たり、動画の生成

に関するユーティリティ機能を提供してもよい。

【0174】その機能の例として、生成すべき中間画像の枚数を指定できてもよい。たとえば、第1画像と第2画像がもともと一秒の間隔をもって撮影されたものであれば、通常はそれらの間に30枚程度のフレームがある。ユーザがGUI等を通して「中間画像30枚」と指定すれば、時間スケールとしては、もとの動画と同じものが得られる。ここで「3000枚」を指定すれば、時間を100倍に引き延ばしたスローモーション画像が得られる。本発明では、この枚数がいくら大きくなっても、単に補間計算のステップ幅を変更すれば済むため、処理にまったく困難な部分はない。現在、一秒間に3000枚のフレームを撮影できるほどのカメラは存在せず、本実施の形態によれば、任意の時間レートのスーパースローモーション画像が得られる。

【0175】実施の形態5

本発明はデータ量を削減して動画を生成できるため、一般にCPUパワー、電池寿命、メモリ容量に制約がありがちなモバイル機器への応用にも有利である。そのため、中間画像の生成を、モバイル機器への適用を前提とした機能のもとで行ってもよい。

【0176】その機能の例として、第1画像または第2画像の一方のみと前記対応情報をもとに生成された中間画像を表示してもよい。その場合、画像をテンポラリにしておくためのメモリの容量が少なく済む。対応情報と第1画像のみで中間画像を生成する場合、第1画像の各画素やメッシュ上のポリゴンその他の画像要素（以下単に画素で代表させる）を対応情報にしたがって移動させればよい。すなわち、通常は対応情報で画素の位置および画素値すなわち色の両方を補間できるが、これを画素の位置の移動のみの処理に簡素化する。

【0177】別の機能として、第1画像のみと対応情報をストリーミング処理して中間画像を生成してもよい。一例として、

1. $t=0$ の画像 $I(0)$ と対応情報から $t=0\sim 1$ の中間画像を生成して表示し、画像 $I(0)$ を捨てる
 2. $t=1$ の画像 $I(1)$ と対応情報から $t=1\sim 2$ の中間画像を生成して表示し、画像 $I(1)$ を捨てる
- という処理を繰り返すことにより、比較的小さなメモリ容量で動画のストリーミング表示が実現する。いずれの場合も、中間画像は表示の瞬間だけフレームバッファに存在すればよい。

【0178】以上、本発明をいくつかの実施の形態をもとに説明した。これらの機能または処理は、いずれもコンピュータプログラムの形で提供することができる。また、以上の処理の任意の組合せや組み替えもまた、本発明の態様として有効である。

【図面の簡単な説明】

【図1】 図1(a)とは図1(b)は、ふたりの人物の顔に平均化フィルタを施して得られる画像、図1

(c)と図1(d)は、ふたりの人物の顔に関して前提技術で求められる $p(5, 0)$ の画像、図1(e)と図1(f)は、ふたりの人物の顔に関して前提技術で求められる $p(5, 1)$ の画像、図1(g)と図1(h)は、ふたりの人物の顔に関して前提技術で求められる $p(5, 2)$ の画像、図1(i)と図1(j)は、ふたりの人物の顔に関して前提技術で求められる $p(5, 3)$ の画像をそれぞれディスプレイ上に表示した中間調画像の写真である。

【図2】 図2(R)はもとの四辺形を示す図、図2(A)、図2(B)、図2(C)、図2(D)、図2(E)はそれぞれ相続四辺形を示す図である。

【図3】 始点画像と終点画像の関係、および第 m レベルと第 $m-1$ レベルの関係を相続四辺形を用いて示す図である。

【図4】 パラメータ η とエネルギー C_f の関係を示す図である。

【図5】 図5(a)、図5(b)は、ある点に関する画像が全単射条件を満たすか否かを外積計算から求める様子を示す図である。

【図6】 前提技術の全体手順を示すフローチャートである。

【図7】 図6のS1の詳細を示すフローチャートである。

【図8】 図7のS10の詳細を示すフローチャートである。

【図9】 第 m レベルの画像の一部と、第 $m-1$ レベルの画像の一部の対応関係を示す図である。

【図10】 前提技術で生成された始点階層画像を示す図である。

【図11】 図6のS2に進む前に、マッチング評価の準備の手順を示す図である。

【図12】 図6のS2の詳細を示すフローチャートである。

【図13】 第0レベルにおいて副写像を決定する様子を示す図である。

【図14】 第1レベルにおいて副写像を決定する様子を示す図である。

【図15】 図12のS21の詳細を示すフローチャートである。

【図16】 ある $f(m, s)$ について λ を変えながら求められた $f(m, s)$ ($\lambda = i \Delta \lambda$) に対応するエネルギー $C(m, s)_f$ の挙動を示す図である。

【図17】 η を変えながら求められた $f(n)$ ($\eta = i \Delta \eta$) ($i = 0, 1, \dots$) に対応するエネルギー $C(n)_f$ の挙動を示す図である。

【図18】 第1画像と第2画像間である画素の対応の様子を示す図である。

【図19】 第1画像上にとられた起点ポリゴンと第2画像上にとられた終点ポリゴンの対応関係を示す図である。

【図20】 起点ポリゴン内の点に対応する終点ポリゴン内の点を求める手順を示す図である。

【図21】 実施の形態1で対応点ファイルを生成する手順を示すフローチャートである。

【図22】 実施の形態1で、対応点ファイルをもとに中間画像を生成する手順を示すフローチャートである。

【図23】 実施の形態1に係る映像効果装置の構成図である。

【符号の説明】

- 10 映像効果装置
- 12 画像データ入力部
- 14 マッチングプロセッサ
- 16 対応点ファイル保持部
- 18 中間画像生成部
- 20 表示部
- 22 通信部

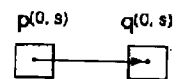
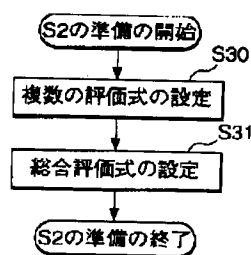
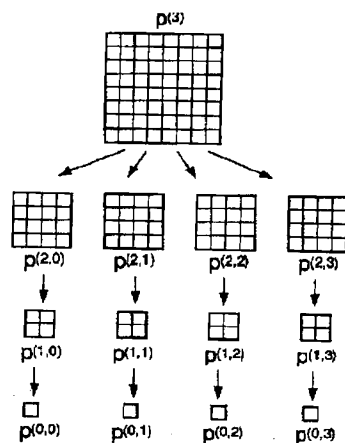
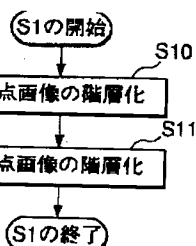
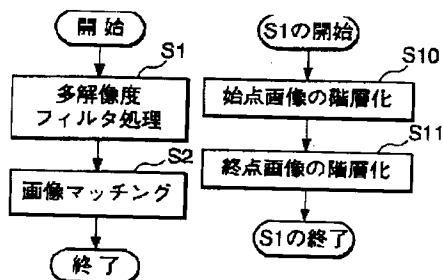
【図6】

【図7】

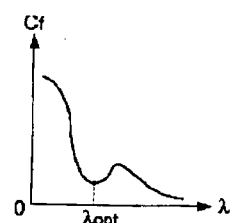
【図10】

【図11】

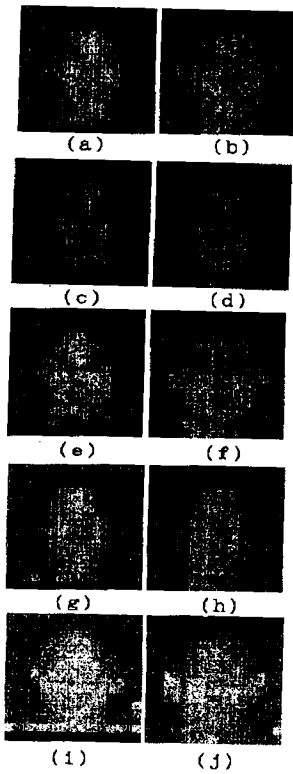
【図13】



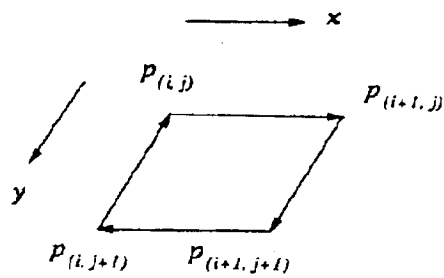
【図16】



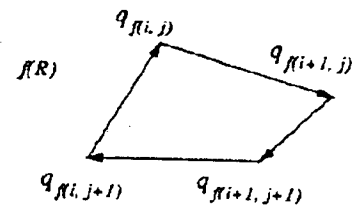
【図1】



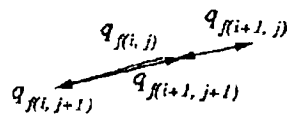
【図2】



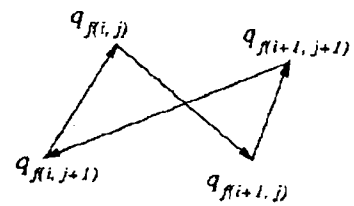
(R)



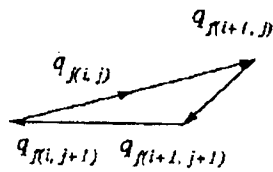
(A)



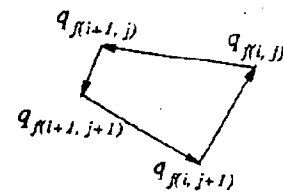
(E)



(B)

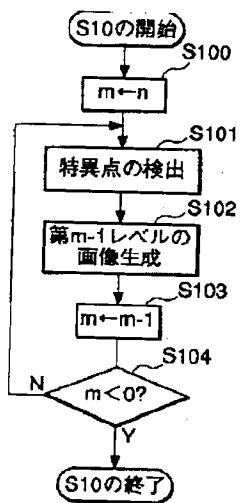


(D)

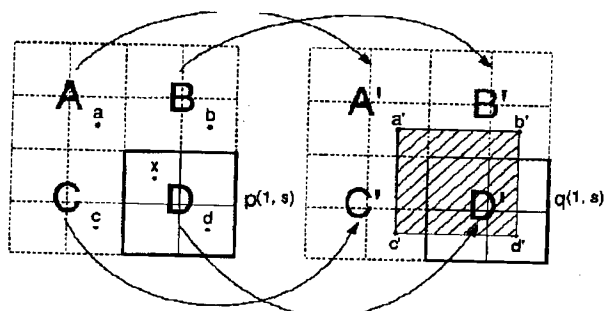


(C)

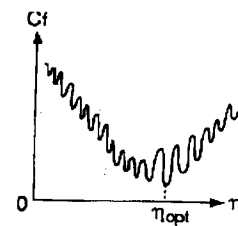
【図8】



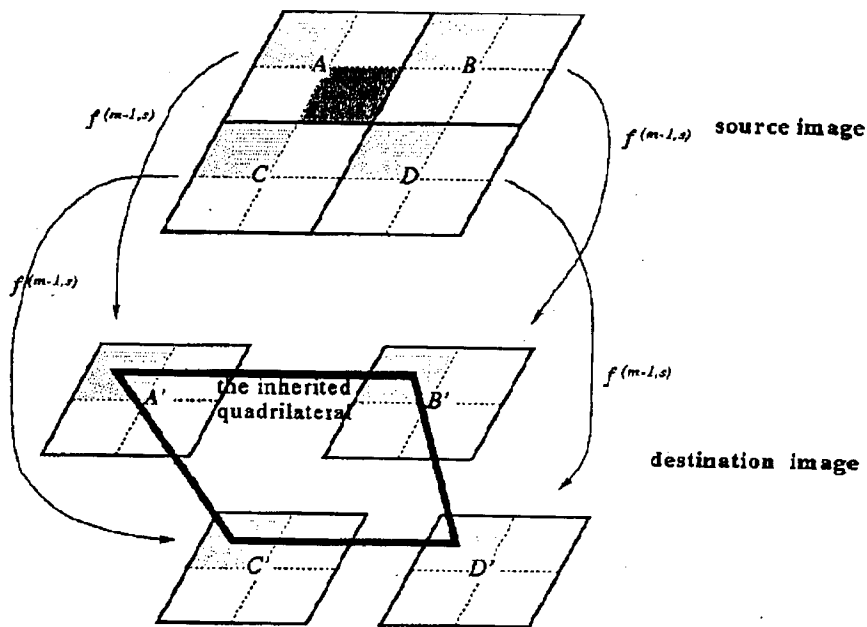
【図14】



【図17】



【図3】



【図9】

$p(m, s)$	
10	8
6	3

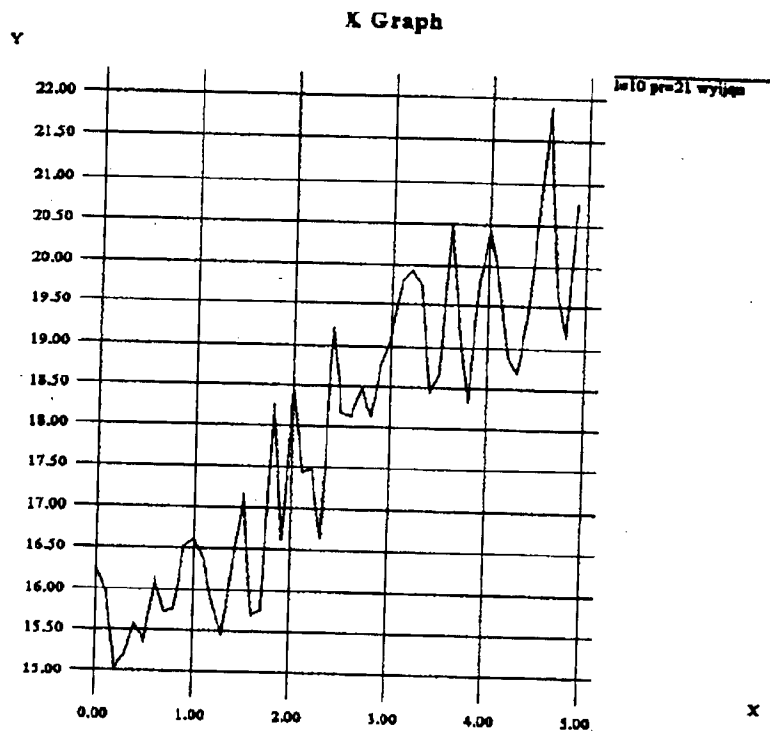
$p(m-1, 0)$	
3	

$p(m-1, 1)$	
8	

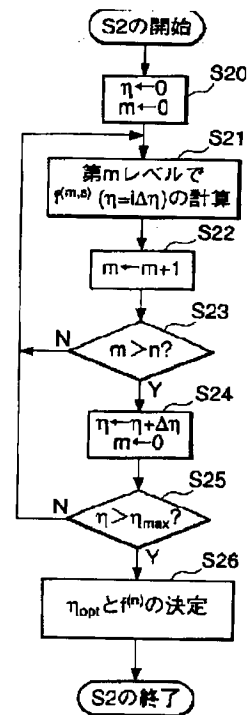
$p(m-1, 2)$	
6	

$p(m-1, 3)$	
10	

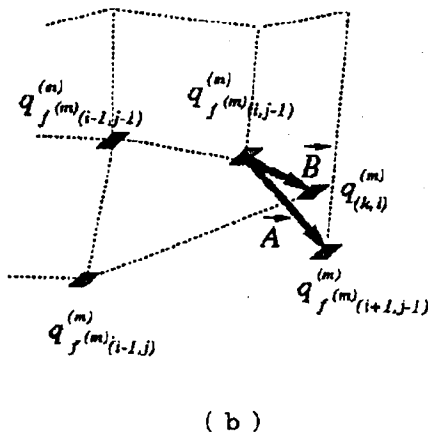
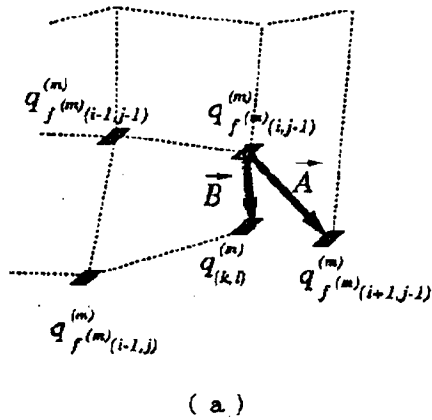
【図4】



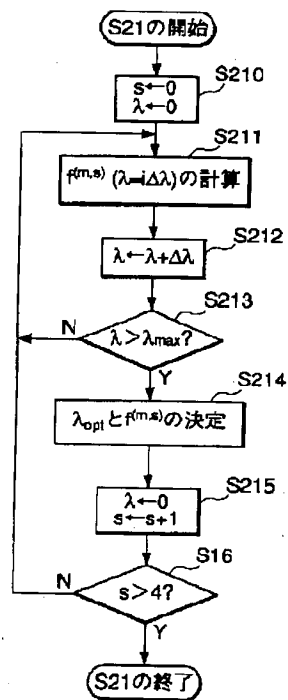
【図12】



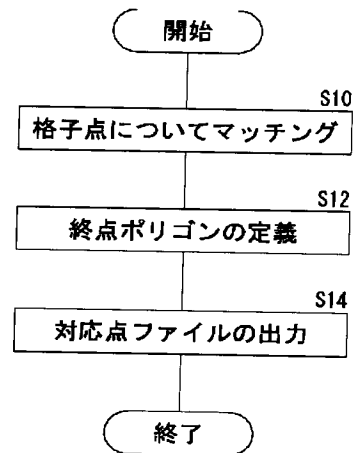
【図5】



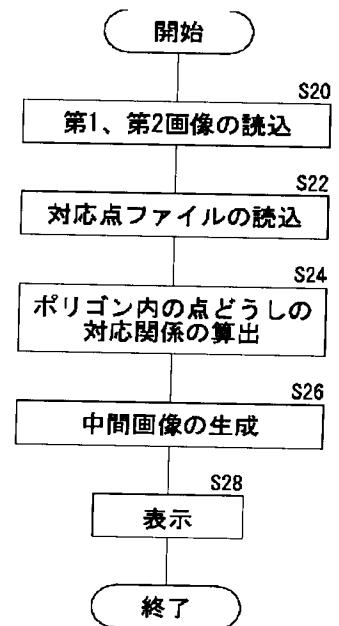
【図15】



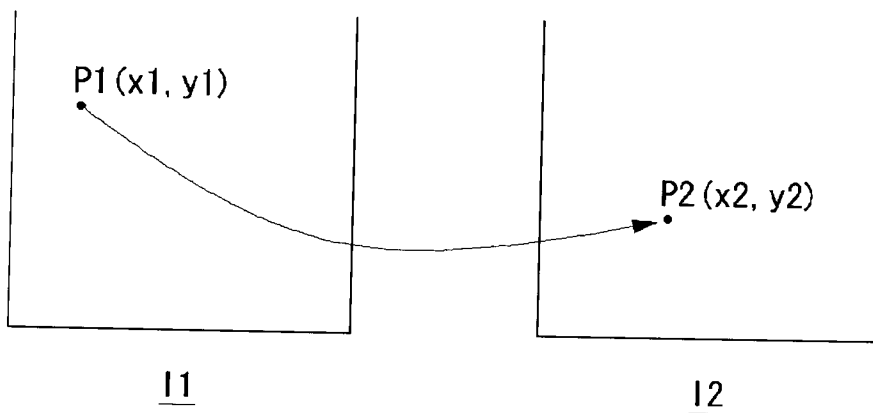
【図21】



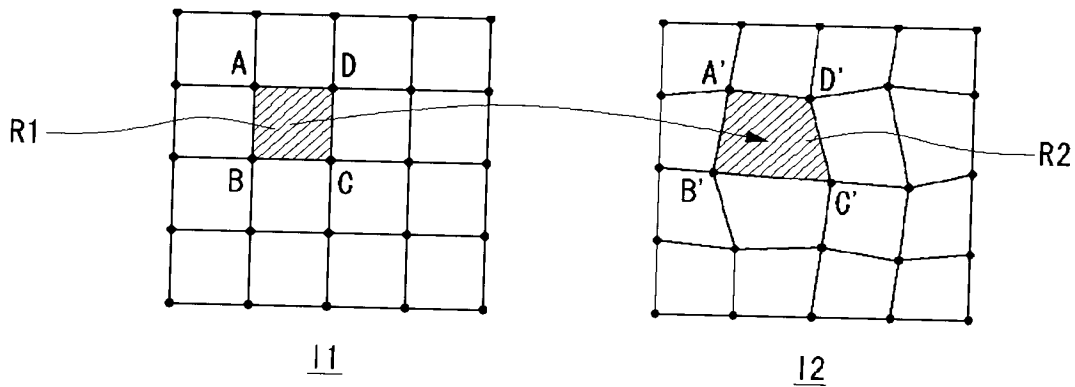
【図22】



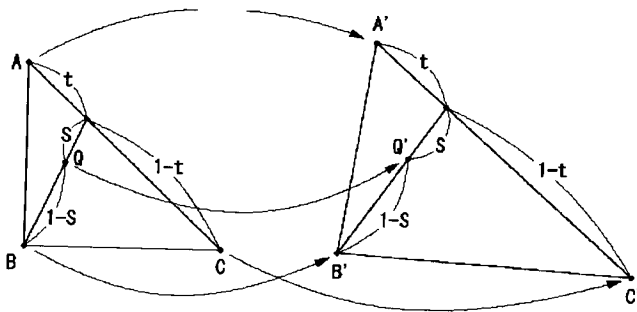
【図18】



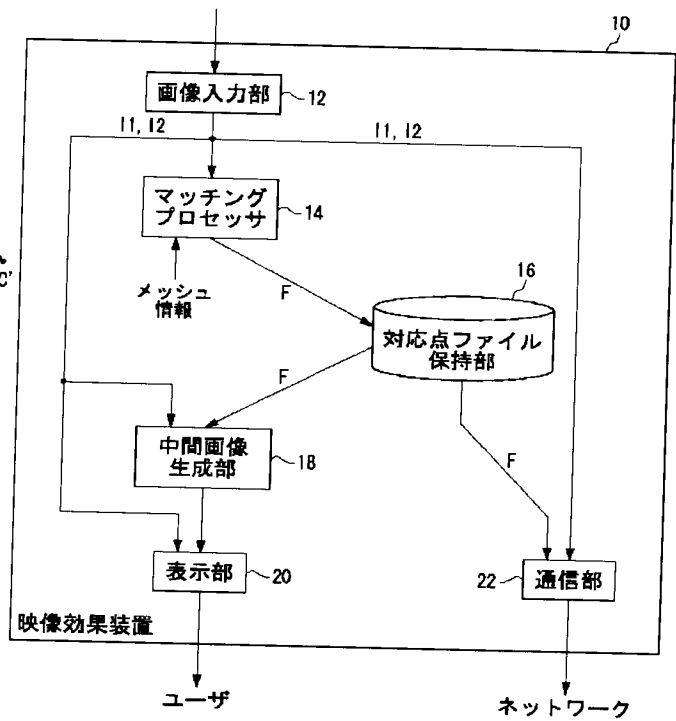
【図19】



【図20】



【図23】



フロントページの続き

(51) Int. Cl. 7

H04N 5/262

識別記号

F I

H04N 5/262

テーマコード (参考)

(72) 発明者 秋吉 信雄

東京都港区麻布十番1-7-3 株式会社
モノリス内

F ターム(参考) 5B050 AA08 BA08 BA09 EA05 EA11
EA24 EA28 FA02
5B057 AA20 BA01 CA12 CB12 CB13
CC01 CD06 CD14 CE06 CE20
CG01 CH01 CH09 CH11 DA16
DC05 DC07
5B080 AA14 BA01 GA21
5C023 AA11 AA38 AA40 BA02 BA11
CA02 CA08 DA04 DA08 EA03
5L096 AA09 BA08 EA33 FA09 FA12
FA13 GA53 GA55 HA02 JA13
JA20